8-21-2019 10:30 AM

Objective and Subjective Evaluation of Binaural Beamformers in Hearing Aids

Scott Aker, *The University of Western Ontario*

Supervisor: Parsa, Vijay, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Engineering Science degree in Electrical and Computer Engineering

Abstract

Hearing aids use a variety of noise reduction techniques to enhance the experience of hearing impaired listeners. One of these techniques is beamforming, which typically aims to preserve sounds coming from the front of the user and suppresses those from the sides and back. Recently, hearing aids have begun employing a wireless connection between the left and right hearing aids in order to augment the directionality of the beamformers, called binaural beamformers. However, the effect of these binaural beamformers on perceived quality and intelligibility has not been thoroughly tested. This thesis investigated the benchmarking of hearing aids which utilize binaural beamforming algorithms using behavioural testing and computational models. Speech recordings from bilateral pairs of several popular hearing aids were obtained across different processing conditions, and in different noisy and reverberant environments. The quality of these recordings was evaluated subjectively by thirteen hearing impaired adults. In addition, computational predictors of perceived quality and intelligibility were extracted from the left and right hearing aid recordings. Objective and subjective analyses revealed that binaural beamforming has a generally positive effect on quality and intelligibility that was dependent on the directionality of the speech and noise. The ear recording with the better predicted quality score was also found to correlate better with the subjective quality ratings than the average of left and right ear predicted scores. A new weighting function that optimally combines the monaural computational metrics was developed, which was shown to be especially effective in environments where speech and/or noise sources are asymmetrically positioned.

Keywords

Hearing aids, binaural beamforming, noise reduction, speech quality, speech intelligibility, quality metrics, HASPI, HASQI.

Summary for Lay Audience

Hearing aids use a variety of signal processing techniques to enhance the experience of hearing-impaired listeners across varied listening environments. One of these techniques is noise reduction, where unwanted signals such as ambient noise and unwanted speech are suppressed while signals such as wanted speech are enhanced. Recently, hearing aids have begun utilizing binaural beamformers, which use a wireless link between the left and right hearing aids in order to amplify signals originating from the front of the user while suppressing those from the sides and back. Effectively, the algorithm utilizes the assumption that the user is looking at what they want to listen to in order to reduce noise. However as binaural beamformers have only been recently developed, the actual benefit the algorithms have on enhancing the quality and intelligibility of speech in noisy conditions is largely unknown. This thesis investigated the benchmarking of hearing aids which utilize binaural beamforming algorithms using both computational models of the auditory system as well as behavioural testing with hearing-impaired listeners. Binaural beamformers were found to have a generally positive effect on the quality and intelligibility of speech, however it largely depended on the directionality of the speech and noise. It was also found that when using a computational model to predict speech quality, the better scoring ear was a better predictor of the behavioural testing results. A new weighting function to combine predicted quality scores for the left and right ears was developed that more heavily weights the better scoring ear.

Acknowledgments

I'd first like to thank my supervisor Dr. Vijay Parsa for his continued support and dedication since way back during my undergraduate when I showed up at his office asking if I could volunteer in his lab. Since then, he has invested time and energy teaching me about signal processing and hearing aids through letting me volunteer in his lab, two undergraduate courses, my fourth year capstone project, a graduate class and of course, agreeing to take me on as a master's student. I couldn't ask for a better supervisor, and his guidance and positivity has helped me through innumerable challenges.

I'd also like to thank the staff, colleagues and friends at the NCA for enabling an incredible work environment. A special thanks goes to Paula Folkeard and Steve Beaulac as well as Dr. Susan Scollie, Dr. Ewan Macpherson, Hasan Saleh and Dr. Jonathan Vaisberg for their unwavering support and advice for both my project and academic career.

Likewise I'd like to thank my colleagues and friends in Western Engineering, many of whom have always had my back through both undergraduate and my master's degree.

A huge thank you goes out to the people of Audioscan including Jim Jonkman and Jonathan Pietrobon for the opportunity to intern there and supporting me through my placement. The experience was invaluable to me, I loved working there, learned a lot and working with the Audioscan team was a highlight of my program.

Finally I'd like to thank my friends and family outside of the University for keeping me sane, including Mom, Dad, and my sister Michelle. Thank you for your continued love and support.

Table of Contents

List of Tables

List of Figures

List of Plates

No table of figures entries found.

List of Appendices

Chapter 1

1 Introduction

Hearing aids incorporate a variety of signal processing techniques in order to assist hearing impaired listeners achieve a comfortable hearing experience. Hearing loss can be complicated: not only will there be a different level of hearing loss at each frequency of sound, the inability to perceive quieter sounds has no effect on the threshold where loud sounds become uncomfortable for the same frequency [1]. In effect, hearing aids must be able to accommodate the user's limited dynamic range in order to afford hearing aid users a comfortable listening experience. The utilization of a smaller range is where condition-specific signal processing algorithms such as noise reduction become useful. The thesis will investigate the effectiveness of binaural beamforming algorithms, a noise reduction technique, in enhancing speech intelligibility and quality in speech in noise conditions.

1.1 Signal Processing in Hearing Aids

Modern hearing aids make heavy use of signal processing, as shown in Figure 1.1, which illustrates the general signal processing techniques a hearing aid applies through from input to output. The figure is separated into sections: Sound Pick Up, where microphones in the hearing aid pick up audio and sort it into an array, Sound Cleaning, where the signal is pre-processed with noise reduction or feedback cancelation algorithms used to ensure as few undesirable aspects of the signal are cleansed as much as possible before being sent to Audibility & Loudness, where the hearing aid applies gain according to the user's audiogram and fulfills the primary purpose of the hearing aid. Throughout this process, Environment Classification processes work to use the input signals to identify the situation the user is using the hearing aids in and adjust the processing, or steer the hearing aid, accordingly. Each of these processes plays an important role in delivering a comfortable and intelligible auditory experience to the user, and likewise, each process has performance measures that can be evaluated in order to provide a clear picture of the performance of the hearing aid on a whole. In the case of noise reduction, the amount of noise before and after the process has taken place can be measured, as can the intelligibility and quality of speech processed by the algorithm.

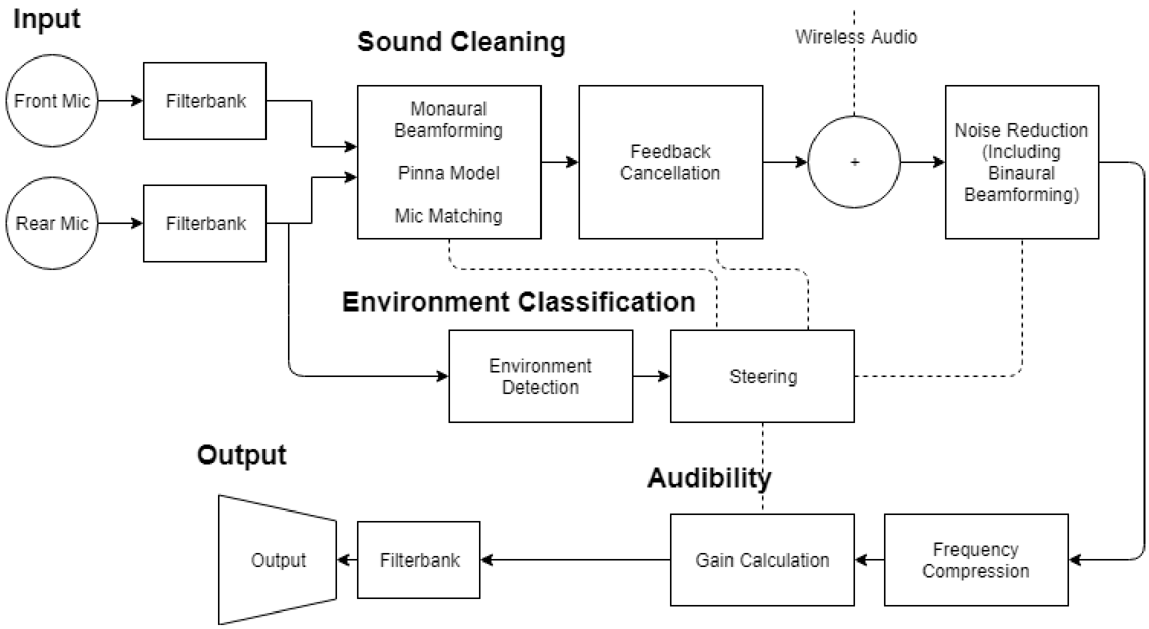


Figure 1.1: Signal processing in hearing aid [2].

1.2 Binaural Beamforming

Noise reduction in hearing aids can be defined as removing unwanted sounds, noise, from a signal while still ensuring wanted sounds remain undistorted. The human brain is remarkably effective at this problem, able to parse out individual voices in crowded, noisy areas in a phenomenon dubbed the “cocktail party phenomenon.” Replicating the same ability with signal processing, however, has proven difficult [3]. Noise reducing binaural beamforming algorithms have recently become a popular solution in modern hearing aids to alleviate this problem. A beamforming algorithm is a signal processing technique which enhances signal from a certain direction while suppressing signals from other directions. In hearing aids, beamformers are used to enhance the signal originating from the front of the user while suppressing noise from the sides and back, effectively amplifying sound sources the user is facing while attenuating sound sources around them. A hearing aid with beamforming algorithms, then, aims to enhance the user’s ability to focus on speech in situations such as a one-on-one conversation while reducing unwanted speech or ambient noise in the background.

As hearing aids continue to evolve, recent beamforming algorithms have begun to utilize the binaural link between left and right hearing aids to narrow the range of angles where the signal is amplified instead of suppressed. Binaural hearing aids, as opposed to monaural hearing aids, are hearing aids which can wirelessly communicate between the left and right devices. Binaural hearing aids can have a range of advantages including easier human interface, as program changes to one hearing aid such as volume control will change the settings in other as well, as well as the ability to transmit the entire audio signal from one hearing aid to the other in cases where the desired signal originates from one side (such as listening in a car or in cases where the user has severe asymmetric hearing loss). Beamformers which take advantage of the wireless link in binaural hearing aids can utilize the transmitted input signal from the contralateral hearing aid for a total of four input microphones, instead of the usual two, which can be used in the beamformer calculations. In Figure 1.1, this can be identified by the Wireless Audio arrow prior to the noise reduction stage. Commercial implementations of binaural beamformers are relatively recent, with Phonak announcing their Quest platform capable of binaural beamforming in 2012 [4] and Siemens introducing a binaural beamforming algorithm in 2014 [5]. While the mathematical operation behind the beamformer can vary depending on the hearing aid manufacturer, the additional input generally leads to a beam that, while still dependent on the beamforming function itself, can be narrower than that achieved by a monaural beamformer [2].

1.3 Assessment of Hearing Aid Features

Speech recordings, and by extension, noise reduction algorithms, can have their performance evaluated in a variety of ways including speech intelligibility and speech quality. Speech intelligibility is a measure of how well the speech can be understood. For instance, playing a recording of speech to a participant and recording how many words the listener correctly repeated back would constitute a simple test of intelligibility. Quality, on the other hand, is a measure of how “good” or pleasant the speech is perceived. A simple speech quality test may involve a listener ranking a speech sample on a scale of zero to five. Speech may be perfectly intelligible, but still maintain annoying or unpleasant distortions that affect its perceived quality.

Testing speech recordings with participants, called a subjective test, is considered the most relevant test for either speech metric. However, behavioural testing can be a costly and time-consuming venture. It is the logistical cost of these tests that led researchers to begin developing ways of predicting subjective testing results with computational models applied to audio recordings. The ability to test for intelligibility and quality without live participants is where the distinction between subjective and objective testing is made, where subjective tests are those which require human participants and objective tests are based on inherent, unchanging features of a hearing aid such as signal processing and computational models.

The hearing aid metrics can therefore be split into four categories, objective and subjective intelligibility tests, and objective and subjective quality tests. Table 1.1 lists examples for each of these categories where the Hearing in Noise Test (HINT) [6] is a subjective sentence intelligibility measure, Multiple Stimuli with Hidden Reference and Anchors (MUSHRA) [7] is the subjective speech quality evaluation method, Hearing Aid Speech Perception Index (HASPI) [8] is the computational predictor of speech intelligibility, and Hearing Aid Speech Quality Index (HASQI) [9], [10] is the complementary objective speech quality predictor. Of these four, objective and subjective quality as well as objective intelligibility were chosen to be the focus for this thesis. Subjective quality testing was firstly deemed to be an area the binaural beamformers had not been thoroughly tested in, and therefore a good choice of experiment to be performed. An objective quality test to compare it to was selected alongside it, and as objective tests are logistically simple to undergo, an objective intelligibility test was selected to be performed as well. MUSHRA, HASPI, and HASQI were ultimately the evaluation methods chosen in those categories, MUSHRA for its statistical validity with a lower number of data points and HASQI and HASPI due to their emphasis on testing the quality and intelligibility of hearing aids specifically. Ultimately, single number indices are derived from each of these scores. In the case of MUSHRA, it is a single score from zero to 100, and for HASPI and HASQI it is two scores, one for each ear, from zero to one.

Table 1.1: Hearing Aid Assessment Tools

	Subjective	**Objective**
Intelligibility	HINT	HASPI
Quality	MUSHRA	HASQI

MUSHRA allows participants to rank several audio samples taken in the same noise conditions but with different noise reduction algorithms on a single screen by adjusting sliders which rank each sample from zero to 100. By having each sample in a single set accessible on a single screen, participants, can replay samples as many times as necessary to get a solid listen and rank for each one. Part of the MUSHRA test is the hidden anchor, which attempts to normalize each screen with each other and prevent minor issues with certain samples from driving the score down beyond what the perceived discrepancy was.

HASPI and HASQI are objective scores which utilize computational models of the auditory system to process the speech in noise audio before applying several signal processing techniques to calculate known predictors of quality and intelligibility. In the case of HASPI, these include cepstral correlation and three-level covariance, and for HASQI these include cepstral correlation, coherence, and long-term changes in the signal spectra. The final indices are derived from a mapping function which takes raw values outputted from the model and maps them to subjective scores. As the mapping function must be calculated with pre-existing data, it is the variability and size of the database that the mapping function is trained on that determines the robustness of the metric. By extension, if certain conditions or features in hearing aids are missing from that database, the metrics may not be valid for scenarios in which those features or conditions are present. Constant validation of the metrics is therefore required to ensure that they are robust enough to be generalized. As binaural beamforming is a relatively new technology, it is one such hearing aid processing strategy which has not been fully validated by HASQI or HASPI.

Another issue arising from the use of HASQI and HASPI is the lack of a single, binaural index representing both the left and right ears. As HASQI and HASPI use audio recordings to measure performance, each ear is given a score independent of each other when the metric is used. In binaural systems, where the left and right ear share information, assigning a single number to the entire binaural system is more perceptually relevant than evaluating each device separately. By mapping the left and right scores to the behavioural data in the same way the HASQI and HASPI mapping functions derive their final indices, a weighting function to determine the final score can be found.

1.4 Problem Statement

With the continued development and proliferation of hearing aids which use binaural beamforming algorithms as a major component of their sound cleaning strategy, methods of measuring and benchmarking their performance become critical in order to be able to accurately compare different models and brands of hearing aids. Testing procedures to compare algorithms are both useful for hearing aid manufacturers, who need logistically simple and cost-effective methods of testing algorithms throughout the development process, as well as audiologists who can use benchmarking data to make more educated decisions when prescribing hearing aids to patients.

1.5 Goals

As binaural beamformers are a relatively new technology, currently no such testing procedure exists. The goals of the thesis will therefore be as follows:

1. Develop a procedure to benchmark binaural beamforming algorithms electroacoustically with HASPI and HASQI and behaviourally with MUSHRA.

2. Validate HASQI with subjective data gathered through behavioural tests.

3. Develop a weighting function to combine left and right HASQI scores into a single, index representative of the subjective data.

1.6 Organization

The organization of this thesis is as follows. In Chapter 2, a review of literature on both objective and subjective quality and intelligibility metrics used for hearing aids is undertaken, as well as a review on the current state of research into binaural beamforming hearing aids. Chapter 3 then focusses on the cloud database developed for the thesis, where hearing aid recordings and their respective objective and subjective test scores are stored in an effort to pilot a long-term hearing aid recording repository for future research. Chapter 4 follows and explores the electroacoustic, or objective testing of the effect of binaural beamforming on predicted speech quality and intelligibility in a wide variety of noise conditions, taking advantage of the low logistical barrier to objective testing. Following that, Chapter 5 seeks to look at the behavioural speech quality testing and examine the relationship between the behavioural results and the electroacoustic results. A weighting function is developed which combines left and right HASQI scores into one, perceptually relevant index. Finally, Chapter 6 concludes the thesis.

Chapter 2

2 Literature Review

As the goals of the thesis comprise of both the measurement of binaural beamforming hearing aids as well as the further development of speech quality and intelligibility indices, it is valuable to understand the state of these technologies based on the current literature. Multiple speech quality and intelligibility metrics exist for signal processing applications such as cell phones, television audio, and radio, however not all are well suited for hearing aids. The metrics that have been modified for use with hearing impaired listeners are in a state of constant development, so the state of these algorithms is important to understand before attempting to advance them further. Additionally, while the effect of binaural beamforming programs in hearing aids have not been extensively studied, other studies investigating their effect on localization and preference can give hints towards the behaviour of these new technologies and help explain the results of the electroacoustic tests.

2.1 Measuring Speech Intelligibility

Several tests exist to measure the intelligibility of speech. Measurements serve to identify how understandable a given passage of speech is, rather than its overall quality or appeal.

The Connected Speech Test (CST) [11] is a speech intelligibility test that seeks to replicate real life scenarios in which the listener will have context to help them understand the content of the speech. The test consists of 48 passages which consist of 10 sentences each. The listener is given a word related to a certain topic, such as “windows,” and then must repeat the following sentences which are related to that word in some way.

Speech Reception Threshold (SRT), though can be predicted objectively [12], has been traditionally measured as a subjective test of intelligibility as well. SRT is a measure which describes the level speech must be presented for a listener to correctly identify the speech contents 50% of the time. SRT is often measured with a Hearing in Noise Test (HINT), developed in [6] to be better suited for predicting speech intelligibility in noisy environments than the CST. The HINT provides a large set of recorded sentences that were constructed to be phonemically balanced with each other for use in subjective testing of intelligibility. Since their development, the HINT sentences have found widespread usage for measuring SRTs or other subjective intelligibility tests where the phoneme-make-up of the test sentences are a concern.

Objective measures of intelligibility were first developed by French and Steinberg at Bell Telephone Laboratories [13], which sought to identify measurable components of a speech sample's intelligibility and combine them into one, quantifiable index from zero (not intelligible) to one (intelligible). In [13], it was concluded that the intelligibility of speech is determined by the sum of audible speech components, which formed the mathematical basis of the Articulation Index (AI). In other words, the AI can be determined by splitting the speech into frequency bands and determining what the proportion of audible speech is within that band [14]. The proportion can be determined by simply subtracting the noise level in decibels from the idealized speech spectrum. Each band is then multiplied by its weighted importance to speech intelligibility and then summed to get a final index. While the original AI was first developed in 1947, the metric was again validated in [15] and adopted by the American National Standards Institute (ANSI) in 1969 (ANSI S3.5). Despite its widespread usage, the AI was focused on frequency domain distortions such as noise and band filtering.

One of the successors to the AI is the Speech Transmission Index (STI), a metric developed in 1980 [16], which extended the method used in the AI to account for non-linear distortions and time domain distortions, which in [17] was proved to correlate with the subjective intelligibility tests in hearing impaired listeners. Despite this, the STI was never formally inducted into any ANSI standard.

The Speech Intelligibility Index (SII) standard, which may be considered the "true" successor to the AI, updates the AI through incorporating additional procedures developed in the STI, to further increase the SII's ability to account for non-linear distortions and time domain distortions such as echo and reverberation [18]. The list of distortions the SII has been updated to account for include fluctuating backgrounds [19]

[20] and binaural listening [21] as discussed later in the chapter. In addition, there exists and extension of SII [22] which includes broadband peak-clipping and center-clipping distortions. This updated metric, termed Coherence-based SII (CSII), simply replaces the SNR estimate parameter of the SII with the speech distortion ratio (SDR) to provide better intelligibility predictions for both normal-hearing and hearing-impaired listeners.

Still, certain processes and distortions often unique to hearing aid processing systems such as ideal time frequency segregation (ITFS), show a low correlation with scores determined by metrics such as CSII and STI [23]. A metric called the Short-Time Objective Intelligibility (STOI) attempts to remedy this [24]. As the name suggest, the STOI works by first performing a Short-Time Fourier Transform (STFT) on both the clean and degraded speech signals, then grouping the resulting time-frequency bands through a third-octave analysis. The data can be visualized as, for each third-octave band, a strip of time-frequency bands which change along the time axis. The "strip" is then split once again into small segments of time, in the case of [24], 384 milliseconds in length. The corresponding segments in the clean and degraded speech are then correlated, and each sample correlation coefficient for each time-octave-band segment is averaged for the final index. The algorithm provided strong correlation with intelligibility scores in ITFS-processed speech signals and signals processed by single-channel noise-reduction algorithms, which were weak points of previous metrics [24].

An attempt to combine the benefits of a coherence metric (such as the CSII) and a short-time envelope metric (such as STOI) was developed in 2014 in the form of the Hearing Aid Speech Perception Index (HASPI) [8], the block diagram of which is shown in Figure 2.1. HASPI uses two sets of raw, objective values which are then mapped to a dataset of intelligibility scores. The first, non-linear index, cepstrum correlation, is similar to the STOI in that it is calculated from the correlation between a processed version of the clean and degraded speech signals. After being processed through a computational model of the auditory system, the envelope of the signal is taken. Then, each sequence is approximated using a set of half-cosine basis functions called cepstrum basis functions, functions which can be thought of as the "building blocks" of speech. As this process is completed for both the clean and degraded speech, the new, approximated signals can then be reconstructed and cross correlated with each other to get the cepstrum correlation.

The second component of HASPI is the auditory coherence term, which itself is split into low-, mid- and high-level coherence components. Both the clean and degraded signal are processed with the same computational model of the auditory system and then split into short-time segments of 16 ms. Segments that correspond to silent portions are ignored, the rest of the segments are sorted into low-, mid- and high-level intensity categories. The short-time segments of the clean and degraded signal are then normalized and cross correlated with each other, with the result being averaged among segments in a like intensity category. This analysis leads to three auditory coherence values for low-, mid-, and high-level intensities. The three auditory coherences and the cepstrum correlation are mapped to a dataset of intelligibility scores to provide the final HASPI index.

Hearing loss is incorporated in the model in several ways. First, the gammatone filter bank that models the inner ear uses filter shapes which replicate the outer hair cell (OHC) damage. OHC damage is also modelled through dynamic range compression, which occurs after the signal is modified by the gammatone filter bank. Inner hair cell (IHC) damage is modelled during this step by attenuating the signal according to the subjects hearing loss. As HASPI incorporates hearing loss in its computational model through the subject’s audiogram, it has clear advantages over other metrics for predicting the impact of hearing aid algorithms on speech intelligibility.

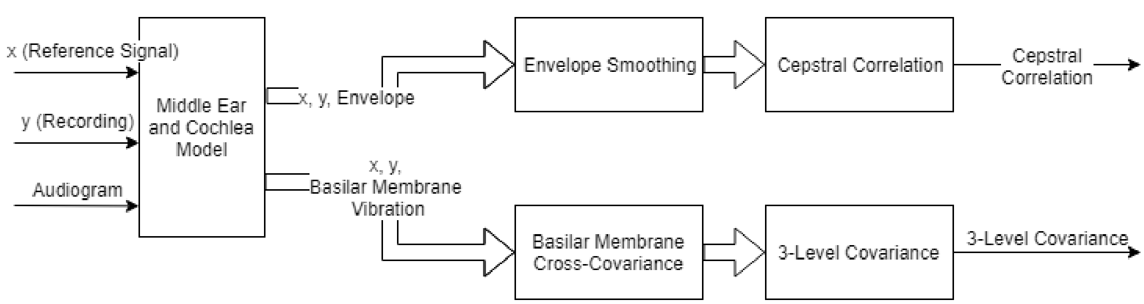


Figure 2.1: Block diagram of HASPI. Output variables are mapped to the final index.

2.2 Measuring Speech Quality

Mean Opinion Score (MOS) is a way of subjectively ranking speech and audio quality into five categories, as described in Table 2.1 below. It is considered as Absolute Category Rating (ACR) subjective evaluation, often used in the telecommunications industry. Many objective metrics also use MOS, particularly those that map features extracted from an audio recording to subjective rankings.

Table 2.1: Mean Opinion Scores

MOS	Quality	Distortion
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly Annoying
2	Poor	Annoying
1	Bad	Very Annoying

Recommended by the International Telecommunication Union – Radio (ITU-R) for “the subjective assessment of intermediate quality levels” is the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) [7] in which participants are presented with several speech samples at once, each representing the output of a different processing algorithm for the same input. The participant can play each speech sample as many times as they wish, and rate each one on a scale from zero to 100. MUSHRA is recommended over MOS due to the increased intra- and inter-rater reliability. As recruiting hearing impaired listeners can be logistically difficult, a subjective assessment methodology that accounts for a relatively low number of participants is advantageous.

The Perceptual Evaluation of Speech Quality (PESQ) [25] and the Perceptual Model – Quality (PEMO-Q) [26], are two objective speech quality metrics developed primarily for the use in the telephone industry. PESQ is the speech quality metric that comes recommended by the ITU Telecommunication Standardization Sector (ITU-T) and is used by phone manufacturers as the industry standard for objective voice quality testing. PESQ uses the degraded signal and its clean version as inputs, and after level aligning, filtering to replicate a phone handset, time aligning, and equalizing, the two signals are put through an auditory transform and then mapped to a prediction of the MOS. As PESQ is still in use today, albeit in an application different from hearing aids, it provides a good baseline metric with proven strengths.

Perceptual Evaluation of Audio Quality (PEAQ) is the complementary standard developed by ITU-R for audio applications. To broaden the applications of PEAQ, PEMO-Q [26] attempts to move away from purely data-driven, score mapping to subjectively tested MOS and towards a more “theoretically sound” computational model of the auditory system to increase robustness. The final PEMO-Q score is derived from a perceptual similarity measure (PSM), and PSMt, a denotation of the fifth percentile of the sequence of instantaneous audio quality. The PSM, as the name implies, is an index from -1 to 1 determined from the correlation coefficient of the reference signal and degraded signal after going through the auditory model. In [26], the PSM alone performed better than PEAQ in most conditions at the cost of higher computational complexity. Since PSM was developed using an auditory model, it can also be used for generic audio quality measures such as music, as opposed to PESQ which was modelled specifically on speech [27].

A speech quality metric designed specifically for hearing aid applications called the Hearing Aid Speech Quality Index (HASQI) was developed by Kates and Arehart in 2010, and then a second version in 2014. Though HASQI predates the development of HASPI, like HASPI, HASQI is derived from two components: a non-linear component, cepstrum correlation, and a linear component based on the long-term spectra of the clean and degraded signal (see Figure 2.2 for HASQI block diagram). Cepstrum correlation is calculated as described in Section 2.1.

The second, linear index is calculated with the long-term spectra of the clean and degraded signals, which is an attempt to capture the effects of linear filters on the speech in the final metric which will go unnoticed by the cepstrum correlation. Essentially, like cepstrum correlation, the signals are processed through a computational model of the auditory system, then their level is averaged over a single utterance. The two signals then have the standard deviation of their spectral difference and spectral slope difference calculated to capture the raw, linear portion of the metric. Spectral difference is simply the difference between the average normalized spectrum levels of the two signals, and spectral slope is the difference of a signals normalized level of a certain gammatone filter index with the previous gammatone filter index.

In 2014, a new version of HASQI was developed which utilizes a new auditory model described in [28] and also developed by Kates.

There are several other objective quality metrics which do not map to subjective scores nor do they use computational models of the auditory system, and therefore may sacrifice accuracy and robustness for lower computational complexity. These metrics include the signal-to-noise ratio enhancement (SNRE), coherence, segmental SNR, log-area ratio, and log-likelihood ratio. These low-computation metrics, as well as PESQ and PSM, were evaluated in [27] on their ability to predict subjective quality scores. The results of the test showed that for noise reduction testing, the SNRE was the superior metric, however perceptual metrics such as PESQ and PSM are better for objective assessments of perceived speech distortion or general quality.

The robustness of HASQI was also evaluated in [29] where it was shown to have similar correlation to MOS as PESQ, log-likelihood ratio, and frequency-weighted segmental SNR. However as hearing aids shape the audio signal based on the user’s audiogram, using a metric that takes this altered frequency spectrum into account is critical, making HASQI a clear choice for the evaluation of binaural beamformers.

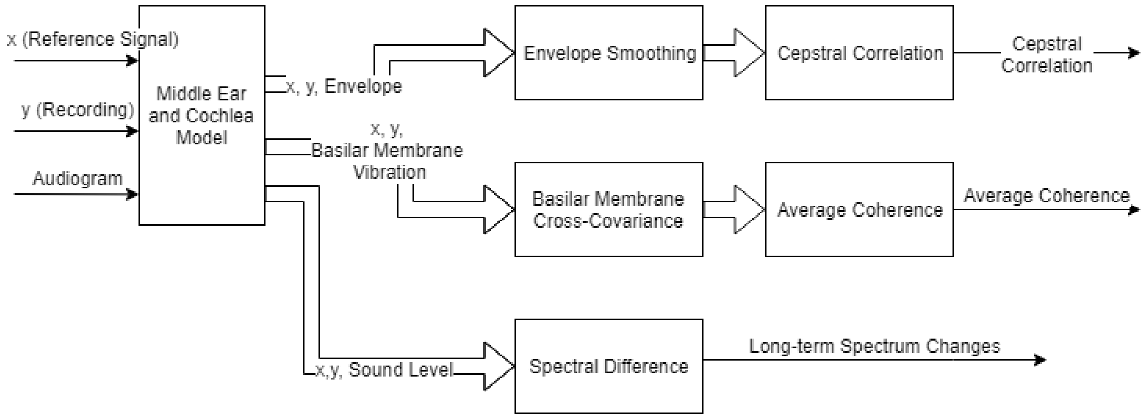


Figure 2.2: Block diagram of HASQI. Output variables are mapped to the final index.

2.3 Binaural Beamforming Evaluation

As testing binaural beamforming in a variety of noise/reverberation conditions is a central component to this thesis, understanding the signal processing behind binaural beamforming was important to interpret the observed behaviour of the beamforming hearing aids with some kind of theoretical foundation. As overviewed in [30], binaural beamforming takes the monaurally processed signal from each ear and wirelessly transfers it to the contralateral ear. It’s notable that the signal being transferred has been monaurally processed with a monaural beamformer, and is not the raw input signal to the microphone. This means by the time the binaural beamforming algorithm is applied, noise from behind the user has already been attenuated.

Binaural beamforming broadly works by adding the two monaurally processed signals from each ear together after appropriately weighting them. As signals originating from the front of the user will be picked up by both hearing aids at approximately the same time, adding the signals together will result in an effect similar to positive interference. A higher weighting will generally be given to the monoaural signal with the minimum power. The reasoning here is that barring the component of the monaural signal that is the same power and phase, which is the signal originating from the front, any additional power is the result of interfering noise. Therefore the signal with the lower power has less noise and more of the 0° signal which is intended to be amplified.

The robustness of specific beamforming algorithms were tested in [31], specifically comparing fixed beamformers versus adaptive beamformers in their susceptibility to steering errors as well as showing the extent to which the test beamforming algorithms distort binaural cues. Part of this analysis, however, included the utilization of objective quality measures to confirm the quality benefit of beamforming in binaural hearing aids. It was this section that was of interest.

Several factors were investigated through the creation of a variety of testing conditions: a total of three binaural algorithms, four head models for use within the algorithms, and the two beamformer types described above (fixed and adaptive). The four head models range in complexity, where the first model simply did not use a head model at all and modelled the sound in free field (FF). Of the two "true" head models, one modelled the head as a sphere (HM1), whereas the second (HM2) used a model developed in [32] which includes near-field and interference effects. Finally, a head-related transfer function (HRTF) was measured on the Brüel & Kjær (B&K) head and torso simulator and used it to model the propagation of sound.

As all the microphones on both the left and right side of the head (XL1-3, XR1-3) contribute to only one binaural output (Z), three strategies were tested to preserve the differences between each side. The first (bin1) involved constructing a filter that was dependent on the output Z, then filtering two input reference signals through it ((XL1, XR1) to get an output for the left and right ears. Second (bin2), was somewhat of the opposite: taking the output Z and filtering it through two propagation vectors (left side and right side) to get the two outputs. Finally, to simulate bilateral hearing aids, the algorithm was simply performed twice for the set of microphones on each side of the head to obtain two outputs.

The objective quality measures chosen included SNRE and PSM, described in Section 2.2 as well as an objective model of SRT proposed in [12]. The beamforming algorithms in bin1 and bin2, which utilize the full set of microphones, performed better than bin3, which was modelled bilaterally, showing the benefit of binaural beamforming in objective quality improvements. However even within bin1 and bin2, different results between the left and right ears for SNRE and SRT required additional explanation. Without a comparison to subjective testing, or a way to combine the left and right quality measures into a single, perceptually representative value, the quantifiable quality benefit of bin1 versus bin2 remained unclear.

Potential subjective benefits from binaural beamforming were discussed in [33], which included testing mild, moderate, and strong levels of directional processing for its effects on localization, sentence recognition, listening effort and participant preference. The “strong” directional processing setting incorporated the binaural beamforming strategy. Localization testing was done through a test called the Spatial Test Requiring Effortful Speech Recognition (STRESR), which had the participant face four loudspeakers at +60, +45, -45, and -60 degree angles, and identify which loudspeaker was playing words. Listeners were then judged on accuracy and reaction time. Localization was found to be negatively affected by beamforming algorithms, as the worst performance was with the strong directional processing. However, with the addition of visual cues, the performance difference between the types of directional processing became negligible, suggesting the differences are “rather small” [33].

Sentence recognition was judged through the Connected Speech Test (CST). The results of the test indicated that performance improved with strong and moderate directional processing, however the only advantage strong directional processing conferred over moderate was in settings with moderate reverberation. Regardless, the benefit of beamforming for speech recognition was shown.

While initial listening effort tests did not reveal a preference between the strong, mild and moderate levels of directional processing, follow up testing showed a stronger preference for strong directional processing over mild or moderate. Additionally, in [34], strong directional processing was shown to improve subjective listening effort as well as objective listening effort in moderate reverberation. It was shown that this improvement might not extrapolate to other reverberation scenarios, suggesting additional testing would be useful.

Beamforming was also tested in [35] to measure the benefits of beamforming algorithms towards solving the “Cocktail Party” problem. The cocktail party problem describes the issues computer algorithms can have isolating speech sources from other spatially separated sources of speech, despite the ease at which the human brain does it subconsciously. The study showed that for listeners with sensorineural hearing loss, beamforming improved the SRT in situations with spatially separated speech-on-speech masking. With that said, for normal-hearing listeners, performance was worse in conditions with beamforming due to the distortion of spatial cues such as interaural time differences.

In [36], binaural beamforming algorithms were specifically tested against monaural beamforming algorithms in commercially available hearings aids in sentence recognition and subjective ratings of perceived work, desire to control the situation, willingness to give up, and tiredness. While both beamformers conferred a benefit in all measurements against the omnidirectional programs, the binaural beamformers only provided a small advantage over the monaural beamformer in sentence recognition and tiredness. Notably, the benefits of binaural beamforming were also found to be independent of noise configuration, where two configurations were used: noise sources at 90° and 270°, and noise sources at 45°, 135°, 225°, and 315°.

2.4 Summary

Common among the literature is that directional noise reduction processing of any kind has advantages over omnidirectional processing, and that binaural beamforming has a small benefit in certain measurements such as speech recognition over monaural beamforming in hearing-impaired listeners. Other findings include:

- The sentence recognition advantages of binaural beamformers were suggested to be dependent on reverberation.
- The speech recognition advantages of binaural beamformers were suggested to be independent to noise configuration.
- Binaural beamformers negatively affect localization, however the negative effect can be tempered by visual cues.

- Binaural beamformers do not have an equal effect on both ears.

Overall, in comparison to published research on the assessment of other hearing aid signal processing algorithms, there is sparse literature on the effectiveness of binaural beamforming algorithms. As such, more reliable and consistent methods of measuring binaural beamforming algorithm performance is a necessity. Additionally, HASQI and HASPI were determined to be the best choice of objective hearing aid assessment for speech quality and intelligibility respectively, but they have yet to be validated for assessing binaural beamforming performance. MUSHRA provides the best methodology for testing speech quality in participants, as it allows for a lower number of participants with higher statistical reliability of their ratings, but it has not been utilized for subjective evaluation of binaural beamforming. This thesis addresses these gaps in the literature, as detailed in the next three Chapters.

Chapter 3

3 REDCap Database

Machine learning has become an increasingly popular modelling technique used in modern computer systems. As machine learning algorithms require to be trained on data sets so that the algorithm in use can find patterns in the records and use those patterns to predict future data, developing methods and infrastructure to acquire and hold large amounts of data are of increasing importance in the modern day. Hearing aid metrics are no exception, as many metrics including HASPI and HASQI utilize mapping functions which were "learned" through training them on subjective data sets. Often, a large, generic data set which can be used to train a model is just as useful as a smaller, more specially designed data set.

As a large amount of data was to be collected in order to examine the relationship between the objective and subjective results of binaural beamforming hearing aids, it was deemed prudent to develop the infrastructure to maintain this data for future studies that may require a generic bank of hearing aid recordings and their associated patient data and subjective ratings. A repository in REDCap – an open source, secure, cloud database – was developed in order to hold the hearing aid recording data and any relevant measurements that may be useful in the future.

3.1 REDCap Overview

The database was created using REDCap, a database manager designed initially by Vanderbilt University for medical research projects. As storing patient data on a server requires ethics approval, using an application designed specifically for medical research made sense to streamline the ethics process as well as creating a greater degree of comfort regarding the security and anonymity of the data. Internally, REDCap is structured like a series of forms that must be designed in advance and filled out for every new data entry in the project. For instance, the REDCap admin must create several fields such as "Patient Number," "Age," and "Gender" and assign them to the project. When a new record is to be added to the database, the user must fill out each of these fields for each record.

In this respect, REDCap is not a particularly flexible database manager, particularly for the secondary goal of creating a database which can be used for future projects that have yet to be fully defined. It is easy to foresee situations in which new fields must be added for new projects that were not predicted when the database was created, or vice versa where not every field is necessary for every record. For example, one user may add a field such as “Noise Direction” to identify which direction the recorded noise originated from. If another user is using the database for an intelligibility test that requires no noise, however, then the field would not be applicable nor make sense to include.

3.2 REDCap Interface

To remedy the inflexibility of the REDCap database manager, a user interface, an example of which is shown in Figure 3.1, was created in C# which would automatically structure the data based on custom made “tags” written in each data field’s notes category. REDCap’s metadata had several fields which defined the form the field entry took, such as label and type, and among these was a “notes” section which could be used for miscellaneous items related to the field. The user interface can connect to the REDCap server, download all the metadata and data in the repository and then subsequently sort it based on which tags it belonged to and the values of certain key fields in the form. For example, if a user wanted to view all audio recordings made through a specific patient’s audiogram, the user could scroll through a list of patient numbers, select the requested one, and a list of each data field for every record made with that patient number entered in the “patient” field would come into view. Additionally, when the patient data came into view, it would be visually segregated based on its “tags.” For instance, data tied to the patient such as age, gender, or audiogram, would be listed under a “Patient Data” header, whereas data tied to a specific recording such as SNR or sound pressure level would be under a “Recording Data” header.

The user interface can update several REDCap forms at once depending on its tag, making updating the data in the repository far quicker and more intuitive. As mentioned above, REDCap stores data on a record-by-record basis, meaning an entire new form must be filled out for each new record entered. If there were two audio recordings for one patient, perhaps one with noise coming from the left and the other with noise coming from behind, there is a need to fill out two records and despite much of the data being redundant such as age and gender. With the user interface, however, multiple records can be updated at once based on the “tag” or category of the data changed. If a field for patient age is updated, for example, and the field is correctly tagged as a “patient” field, the interface can easily go through every record in the database and update every record with the same patient number as the one selected.

Figure 3.1 displays the interface prior to connecting to the REDCap server. In order to connect, the Connect button located at the top left of Figure 3.1 must be pressed, which then automatically pulls the data currently stored on the server into the interface as seen in Figure 3.2. From there, a Study can be selected in the top left list box in Figure 3.2. When a given study is selected, the participants associated with that study then populate the Participant list box directly below. Likewise, when a participant is selected, each recording associated with that study and participant combination is displayed in the Recording list box as seen in Figure 3.3.

Data associated with the participant is then displayed under the Participant header and data associated with the recording is then displayed under the Recording ID header as seen in Figure 3.3. In order to decrease clutter, certain data fields, such as Audiogram, have their own sub-fields which can be viewed by clicking the View button which then opens the window seen in Figure 3.4. In order to edit data fields, the check box to the left of the data field name must be checked. The corresponding field can then be edited, and by pressing the “Update Recording” or “Update Participant” button, the REDCap server will be updated with the new information. An example of changing the SNR data field can be seen in Figure 3.5.

In order to add new data fields, the “Edit Data Field” button can be pressed for either participant data or recording data which opens the window seen in Figure 3.6. New fields can then be added and the type, be it a text box or drop down menu, can be specified. By clicking the “Save & Close” button, the REDCap metadata will be updated to include the new data field as seen in Figure 3.7. Adding new fields does not immediately affect other recordings or participants. As seen in Figure 3.8, while the new data fields will be available for other recordings, unless the check box is checked they are not included in that recording’s or participant’s data set.

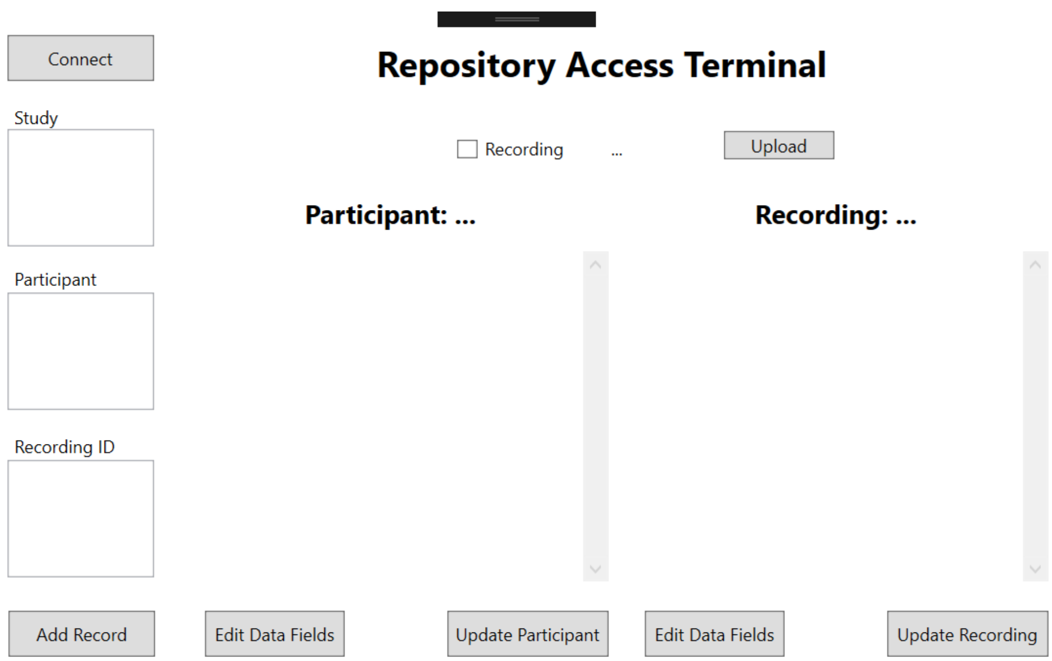


Figure 3.1: REDCap interface prior to connecting to the server.

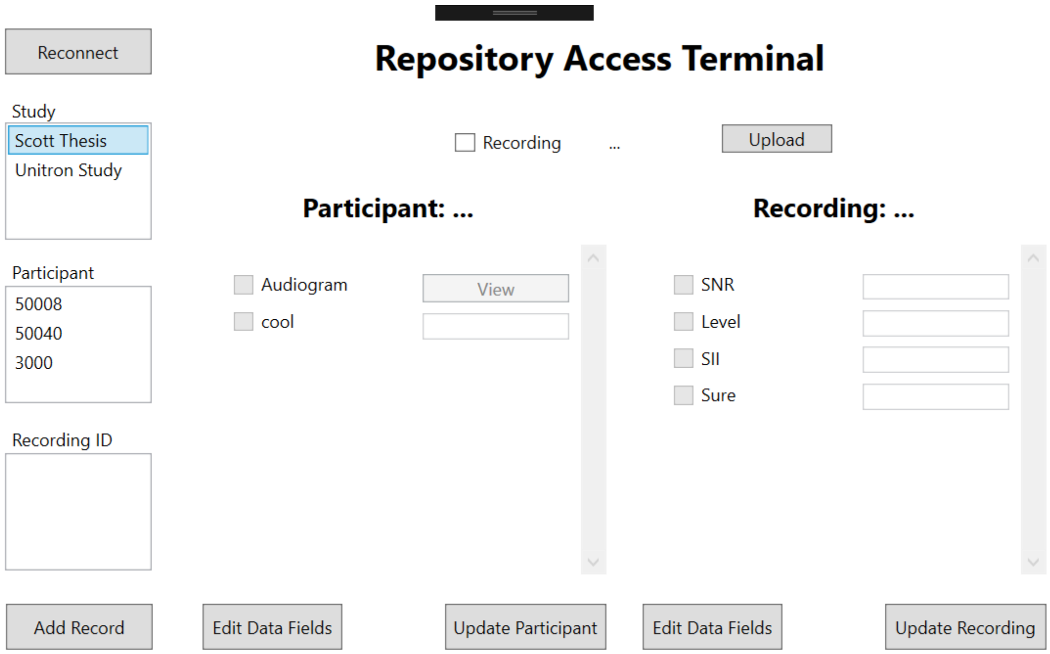


Figure 3.2: Selecting a study in the REDCap Interface.

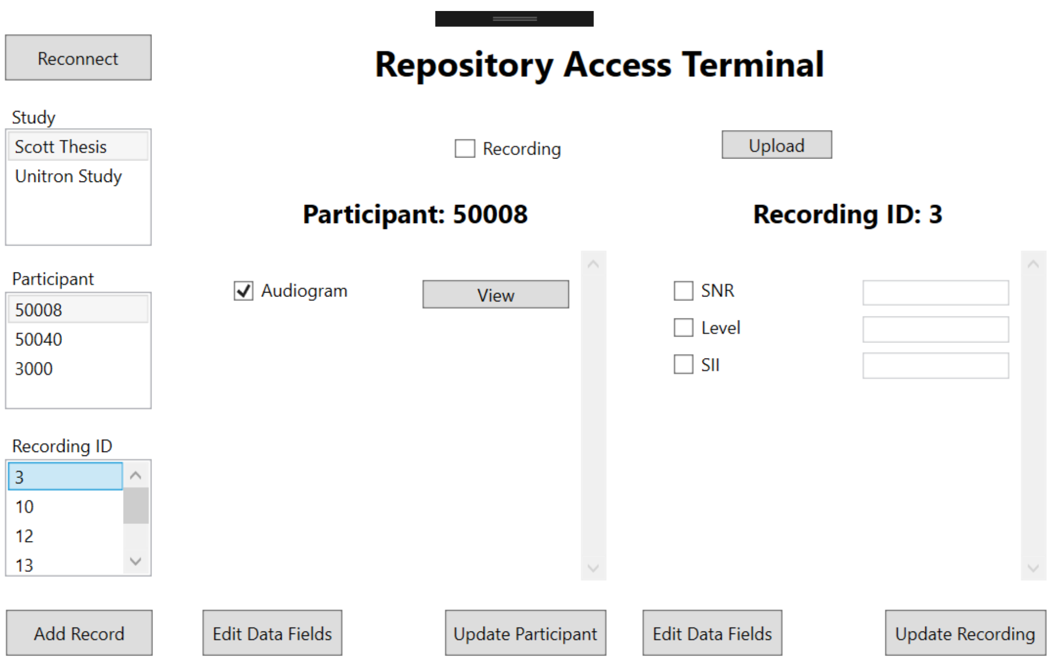


Figure 3.3: Selecting a recording in REDCap interface.

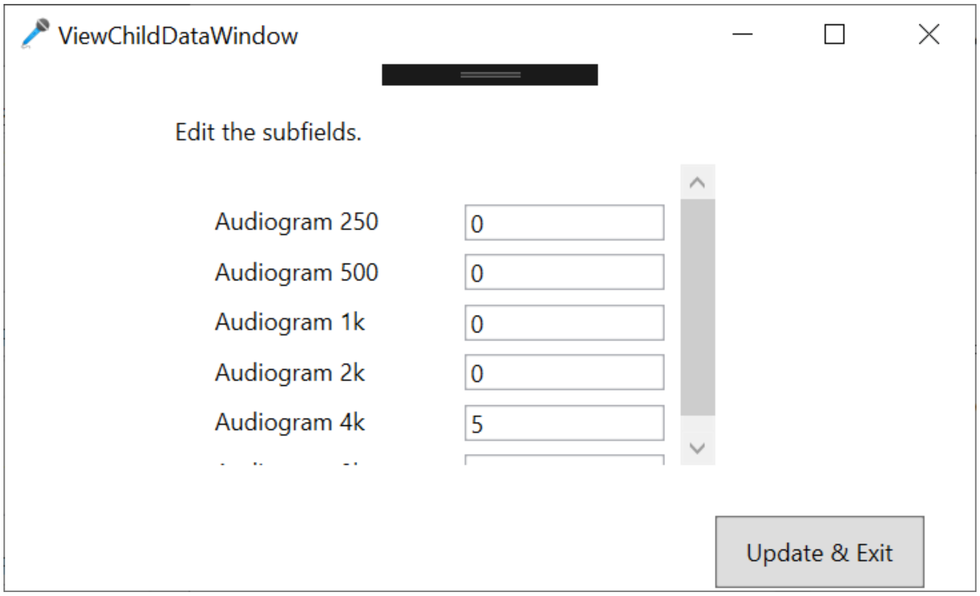


Figure 3.4: Certain data fields with sub-fields have their own interface to decrease clutter.

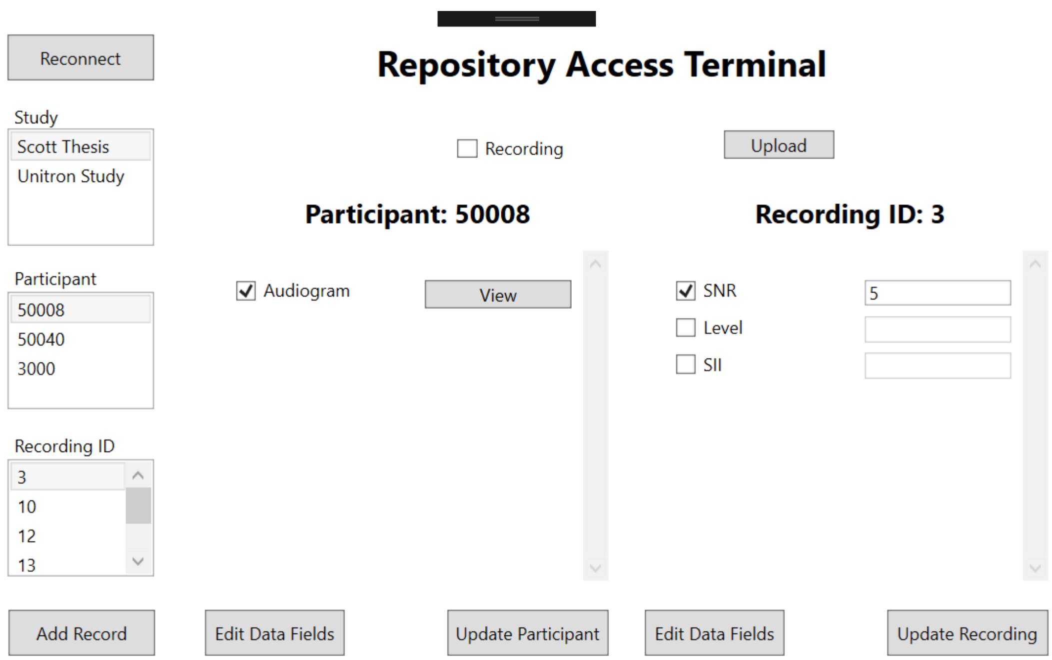


Figure 3.5: Updating the SNR field of a recording in the REDCap interface.

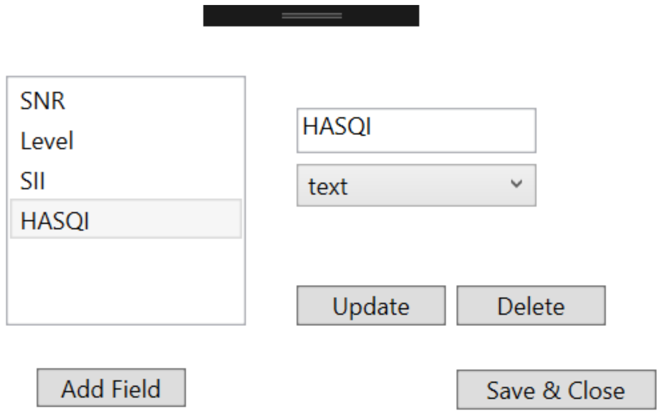


Figure 3.6: Data fields for recordings can be added and edited.

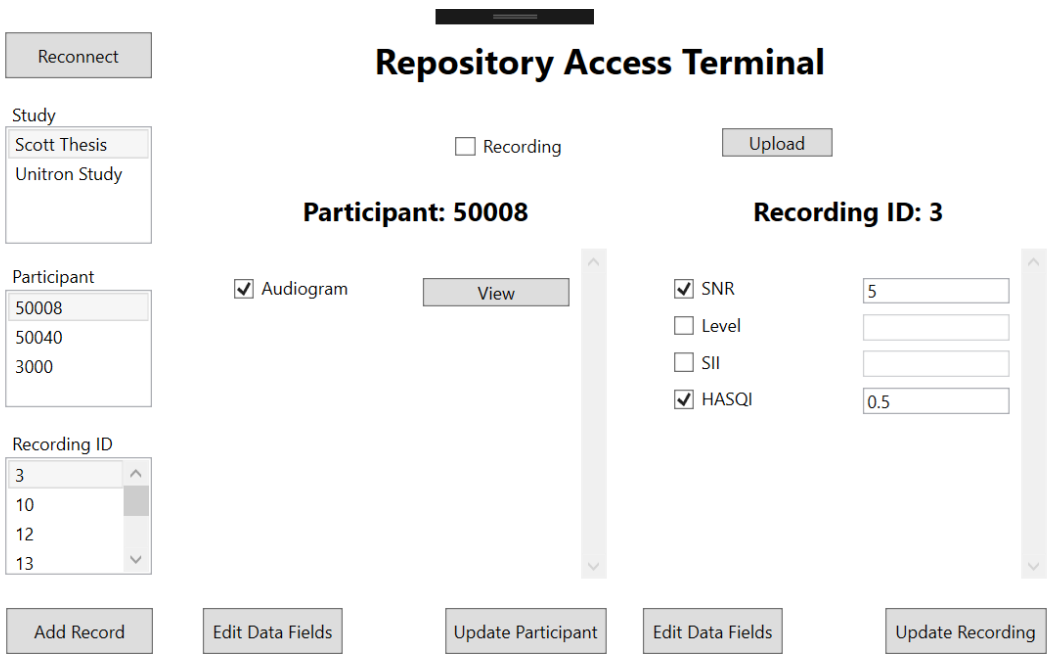


Figure 3.7: REDCap interface with new data field added.

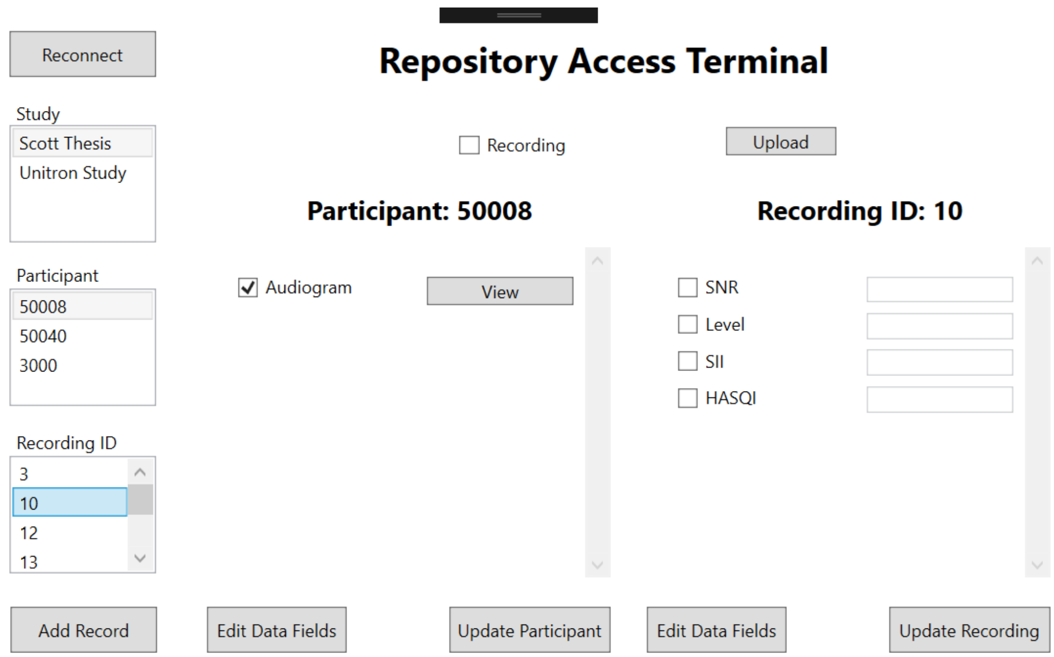


Figure 3.8: Added data fields are available to other recording IDs, but do not affect the database unless the corresponding checkbox is checked.

3.3 Data Fields

Once the internal structure of the database was established, the specific data fields to be used in the project could then be decided upon. Fields were separated into two categories: Participant Fields, which would update every entry under a certain participant number when changed, and Recording Fields, which would only update the specific entry that was changed.

3.3.1 Participant Fields

Age: The age range of the participant was useful when looking at the diversity of the participant sample, and could be recorded while still keeping the participant's identity anonymous.

Audiogram: The audiogram of the participant. As the audiogram of the participant is an integral part of the recording as well as necessary to properly use HASQI, a field to record the hearing loss at each audiometric frequency was imperative.

Years of Hearing Aid Experience: While all participants were experienced hearing aid users, since the years of hearing aid experience was known it was included to parse the data more easily in the future.

3.3.2 Recording Fields

Direction (Noise): The directions of the noise sources where 0° is in front of the user and then rotating around the HATS clockwise.

Direction (Speech): The direction of the speech source where 0° is in front of the user and then rotating around the HATS clockwise.

HASPI: The HASPI score of the recording measured against the clean recording. As HASPI is computationally intensive and can take a long time, uploading the calculated index was done so the HASPI process would not have to be repeated for every new project.

HASQI: The HASQI score of the recording measured against the clean recording. See HASPI.

HASQI CC: The cepstral correlation score used in HASQI of the recording measured against the clean recording. See HASPI.

Level (Speech): The level in dB SPL of the speech measured at the center of the HATS.

Noise Type: The type of noise used. In the case of this study, pink noise, speech-shaped noise and cafeteria noise were all used at some point.

Recording (Clean): A wav file of the original speech sample before being recording through the hearing aid. The clean recording was included as a raw form of the HASQI index in case the HASQI or HASPI results needed to be reproduced.

Recording: A file field to upload the hearing aid recording file. Hearing aid recordings were stored as two channel wav files at a sample rate of 48000 Hz.

Sample Rate: The sample rate of the recording, despite being encoded in the wav file having the sample rate explicitly recording in the database would help parse the data for future projects.

SNR (A priori): The signal-to-noise ratio determined by the level of sound at the center of the HATS.

3.4 Summary

The hearing aid recording repository was developed as part of a larger effort to maintain hearing aid recordings and their associated data from study to study. REDCap was chosen as the database manager due to its common usage for academic studies which often have stringent privacy and ethics restrictions. The lack of flexibility within REDCap was remedied through a database interface developed in C#, which allowed multiple database entries in REDCap to be changed at once if they shared common parameters such as participant number. Data collected through this study was uploaded to the REDCap server to serve as a starting point for the database.

Chapter 4

4 Electroacoustic Analysis

The electroacoustic analysis of the binaural beamforming algorithms in different noise conditions was the first to be performed. Ideally, the audio output of each hearing aid would be captured in a variety of simulated environments, which would then be subject to listening tests by human participants as well as quality and intelligibility metrics such as HASQI and HASPI. However, as there was a more limited number of conditions a participant could reasonably listen to and evaluate in a single session, an initial benchmarking of the hearing aids was performed first, to determine which speech in noise conditions would provide the greatest variety of results as well as to quantify the effect different noise conditions had on the predicted speech quality and intelligibility of the hearing aids.

4.1 Methods

Initial benchmarking recordings were done in two physical environments, the sound booth in the digital signal processing laboratory, and the reverberation chamber, both in the National Centre for Audiology (NCA). The sound booth had a reverberation time of 100 ms, while the reverberation chamber had a reverberation time of 900 ms. Within the sound booth, a B&K Head and Torso Simulator (HATS) sat on a small wooden table, flanked by three loudspeakers affixed to arms which suspended them from the aforementioned table, as depicted in Figure 4.1. The loudspeaker arms were, at their base, attached to a rotating platform on top of the small table. The platform allowed the arms to be rotated about the table, meaning sound could be directed from any of three directions during recording.

The HATS made use of a rubber pinnae to simulate the shape of the outer ear of a patient, and allowed for easy and realistic placement of the tested hearing aids on the manikin, as shown in Figure 4.2. Within the left and right ear canals, microphones led down through torso of the HATS and into a B&K Nexus Conditioning Amplifier which amplified the stereo signals at 100mV/Pa. The signals were then processed with the Echo AudioFire 12 sound card outside the sound booth and recorded with MATLAB on the corresponding computer with the Data Acquisition Toolbox (Figure 4.3). Similar to the hardware involved in the recording, the playback utilized the Echo AudioFire 12 sound card to the AMCRON D-75 multichannel amplifier, then to each corresponding loudspeaker around the HATS.

The computer in the reverberation chamber was also connected to an Echo AudioFire 12 sound card which then fed into a SoundWeb 9088i Networked Signal Processor, which allowed the AudioFire to connect to up to 16 output speakers instead of the usual 8. Finally, the system was connected to a LabGruppen C 10:8X amplifier before connecting to the loudspeaker array within the chamber.

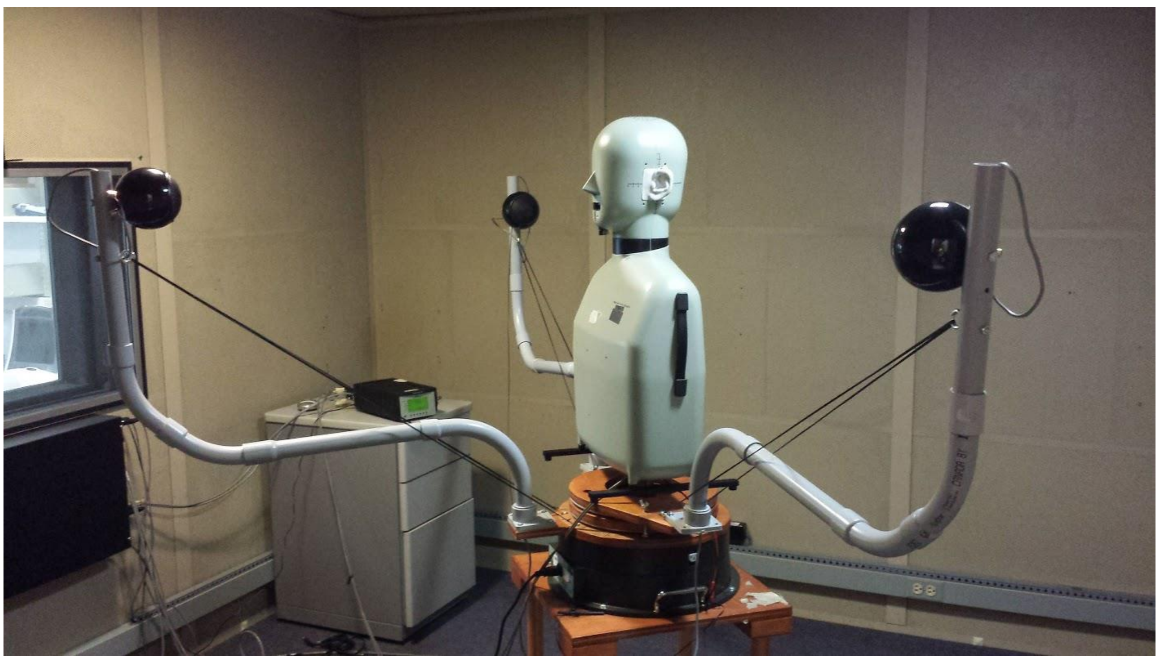


Figure 4.1: B&K HATS on wooden table with rotating speaker apparatus.

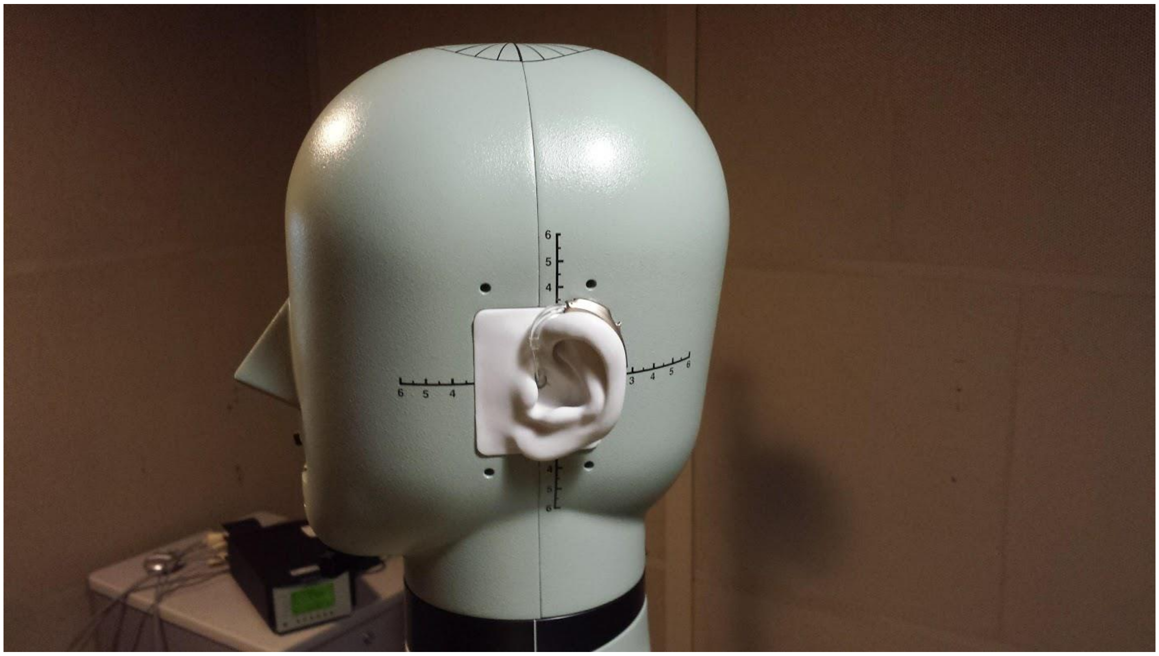


Figure 4.2: Close-up of hearing aid affixed to the rubber ear, simulating the shape and material of a real ear.

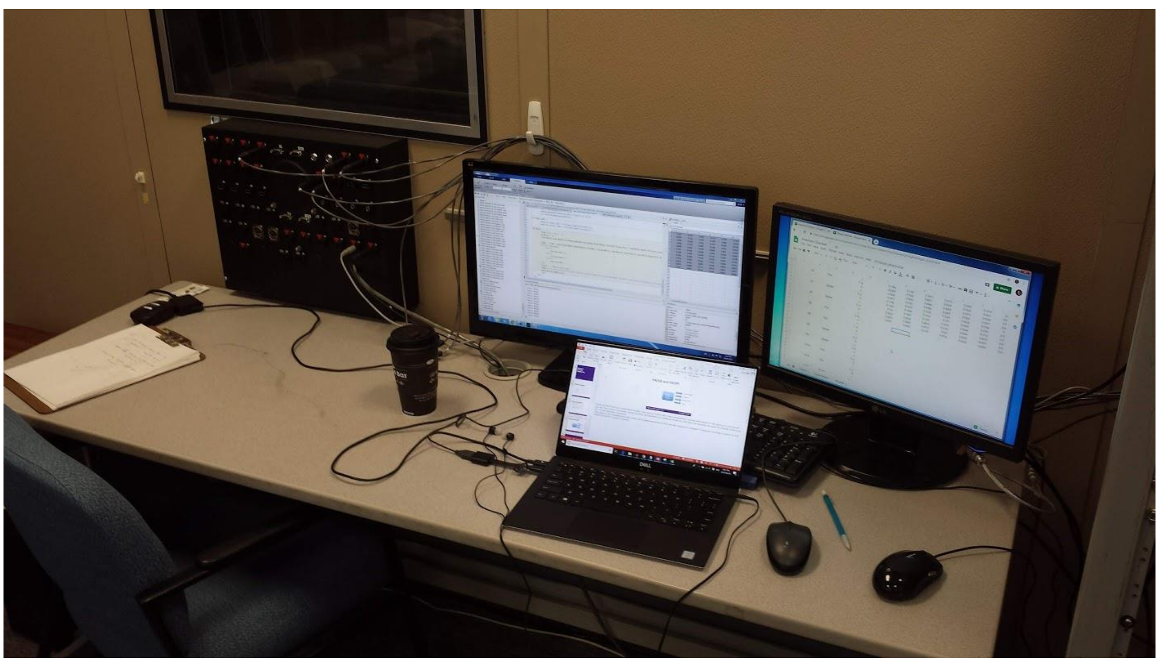


Figure 4.3: Computer set-up outside sound booth used to control speakers and microphones.

While there are a multitude of audio recording methods in MATLAB, the Data Acquisition Toolbox was chosen for its focus on simultaneous playback and recording. Once the playback system was calibrated with a G.R.A.S. Type 26 AK free-field microphone temporarily replacing the HATS at the center of the room and a B&K UA 1546 calibrator which emitted a tone at a known 94 dB sound pressure level (SPL), the precise sound pressure level at the free field microphone could be measured. By then generating pink noise out of each of the three loudspeakers one-by-one, the correct level adjustments could be made to the digital signal outputted at the computer, to ensure not only that all three loudspeakers produced equal sound levels at the center of the room, but that it was at a known sound pressure level which could be calculated based on the digital output level. The B&K UA 1546 calibrator was also used with the HATS in place on each of its ear canal microphones, as being able to calculate the sound pressure level at the ear canal based on the digital input recordings. The recorded SPLs at left and right ears was critical for both HASPI and HASQI measurements, as they incorporate hearing loss model.

4.1.1 Speech and Noise Conditions

To collect a database of the recordings, several parameters of the recording set-ups were adjusted to simulate a variety of noise conditions. Each hearing aid would then be recorded in each condition, on each of the hearing aids' available program settings. Four brands of hearing aids were tested, and most of the hearing aids were programmed to three settings , meant to be switched between by the user depending on their situation or current needs: omnidirectional, which incorporated no noise reduction, a monaural beamforming program, and a binaural beamforming program. All hearing aids were fit to targets prescribed the DSL 5.0 algorithm for the standard N4 audiogram [37].

The direction of the noise itself was also altered between three states: the 90° and 270° state, where the two loudspeakers assigned to output noise would be rotated to the HATS left and right flank; the 90° and 180° state, where one loudspeaker would be to the HATS' direct right and one loudspeaker from its behind; and finally the 45° and 315° state, where the output noise would originate from a point more adjacent to the 0° loudspeaker. The loudspeaker that would play speech would always be at the 0° angle, directly facing the HATS. The speech was initially always played from the front to best utilize the hearing aids binaural beamforming abilities, which amplify sounds from the front while attenuating those from the sides and back. In order to test the beamformers in sub-optimal conditions, the speech was also rotated to a 45° angle when the noise was in its 90° and 270° state, adding a fourth speech/noise spatial configuration condition.

Speech was played at 70 dB SPL, where the sound pressure level was measured at the center of the HATS as described in the calibration process. The noise was played at two different levels for an SNR of 0 and 5 dB respectively. Since there were two loudspeakers dedicated to noise, the noise level from each loudspeaker was reduced by an additional 3 decibels for an SPL of 67 and 62 depending on the desired SNR.

Two noise types were also used: pink noise and cafeteria noise. Pink noise is spectrally and statistically stable while cafeteria noise is non-stationary and meant to resemble the ambient background noise of a restaurant or crowded area, therefore the two noise types provided a spread of realistic noise types.

With four hearing aids, two noise types, two SNRs, two rooms, four directionality conditions, and two to three programs per hearing aid, the described parameters amounted to a total of 352 conditions, which are described in Table 4.1.

The speech samples played were twenty HINT sentences concatenated into one continuous string. Analysis was done only on the last ten sentences in order to provide the hearing aid with at least twenty seconds of settling time in the acoustic environment. Each of the last ten sentences were analyzed independently, then averaged together to get the final result.

Table 4.1: Hearing aid recording conditions for electroacoustic measurements.

Speech and Noise Directions	**Hearing Aids**	**Programs**	**Noise Types**	**SNR**	**Rooms**
• Speech at 0°, Noise and 90° and 270° • Speech at 0°, Noise at 45° and 315° • Speech at 0°, Noise at 90° and 180° • Speech at 45°, Noise at 90° and 270°	• Hearing Aid 1 • Hearing Aid 2 • Hearing Aid 3 • Hearing Aid 4	• Omnidirectional • Monaural Beamformer • Binaural Beamformer*	• Pink • Cafeteria	• 0 dB • 5 dB	• Sound booth • Reverb Chamber

*Binaural beamforming programs were not available on Hearing Aid 4

4.2 Electroacoustic Analysis Results

4.2.1 Sound Booth

The HASPI and HASQI scores of speech in pink noise at 0 dB SNR with the speech originating from 0° and noise from 90° and 270° recorded in the anechoic sound booth are displayed in Figure 4.4 and Figure 4.5 respectively, and provide a baseline for the other measurements due to its low reverberation, statistically flat noise, and speech originating from directly in front of the HATS where binaural beamformers are most optimized to listen to speech from.

When each of the hearing aids are in their omnidirectional program (Omni) with no noise reduction, the HASPI and HASQI scores are similar between different brands. This follows logically as other than adjusting the signal according to the amplification requirements associated with the N4 audiogram, the hearing aids do not provide any other processing in this program leaving little room for deviation between brands.

A larger difference occurs when the hearing aids were switched to program 2, the monaural beamformer (BF), which sees a large improvement in both HASPI and HASQI scores jumping from an average HASPI score of 0.035 to 0.34 and an average HASQI score of 0.059 to 0.24. With the introduction of noise reduction, not only does the processed signal improve in predicted intelligibility and quality, but more variability occurs between brands as the strengths and weaknesses of different processing strategies within each model are divulged. In the monaural beamforming program, Hearing Aid 2 (HA2) has the best HASPI score at a left-right average of 0.52 while Hearing Aid 1 (HA1) has the best HASQI score at a left-right average of 0.27. The pattern of HASPI and HASQI scores do not always align perfectly with each other. While a higher predicted intelligibility may correspond to an increase in predicted quality, it may also telegraph more noise reduction processing in the hearing aid which can often increase quality-degrading distortions in the speech.

The binaural beamforming program (BBF) for Hearing Aids 1, 2, and 3 see a slight improvement in predicted intelligibility and a slight improvement in predicted quality in Hearing Aids 1 and 3. The lack of significant improvement is not a reflection of poor performance on the part of the binaural beamformers, rather it is simply not a condition in which a narrower beamformer accrues any benefit beyond what the monaural beamformers can already achieve.

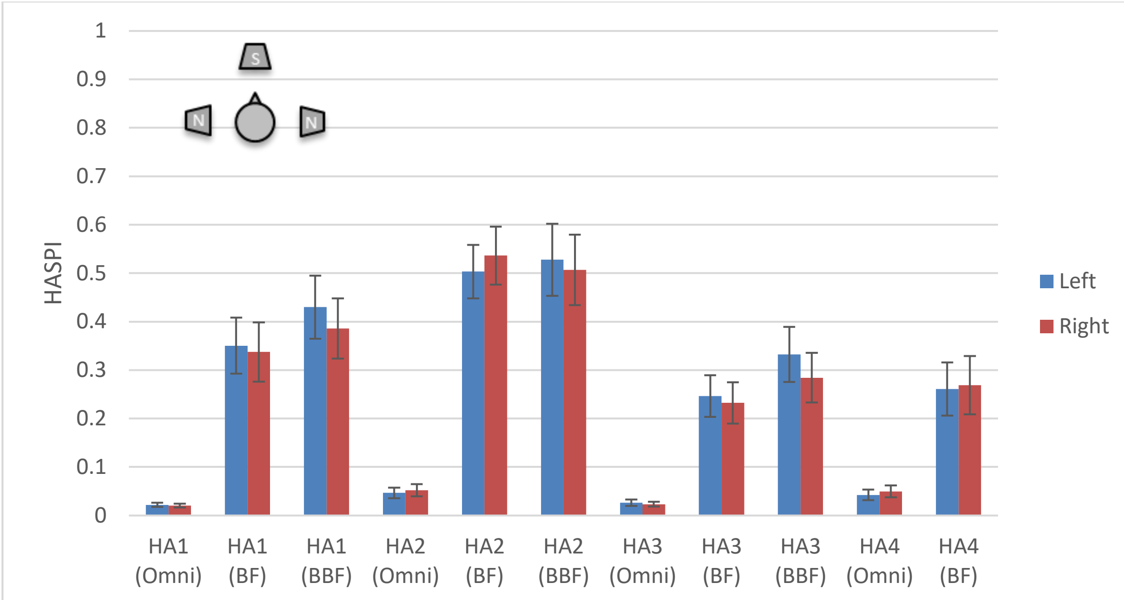


Figure 4.4: HASPI of speech in pink noise at 0 dB SNR, with speech originating from 0° and the noise from 90° and 270° recorded in the sound booth.

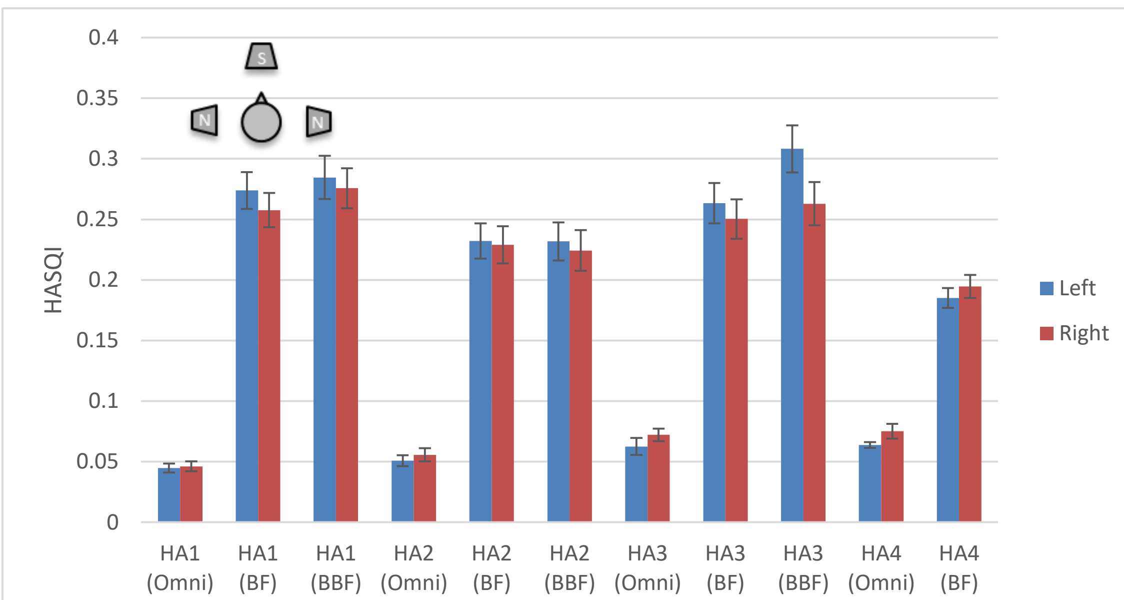


Figure 4.5: HASQI of speech in pink noise at 0 dB SNR, with speech originating from 0° and the noise from 90° and 270° recorded in the sound booth.

The conclusion is further evidenced by Figure 4.6 and Figure 4.7, which show the same pink noise at 0 dB SNR condition however with the noise source originating from 45°

and 315° instead of 90° and 270°. In this case, while the overall HASPI and HASQI scores are smaller than when the noise originates from 90° and 270°, there is a clear improvement from the scores of the monaural beamformers to the binaural beamformers in Hearing Aids 1 and 2. This suggests the benefits of binaural beamformers versus monaural beamformers are best seen when the direction of the noise source are closer to speech at 0° azimuth. As the binaural hearing aids' ability to wirelessly communicate with each other allows for a narrower range of angles in which sound sources are amplified, the narrower beamformer cutting out more distorting noise sources affirms the increase in predicted performance.

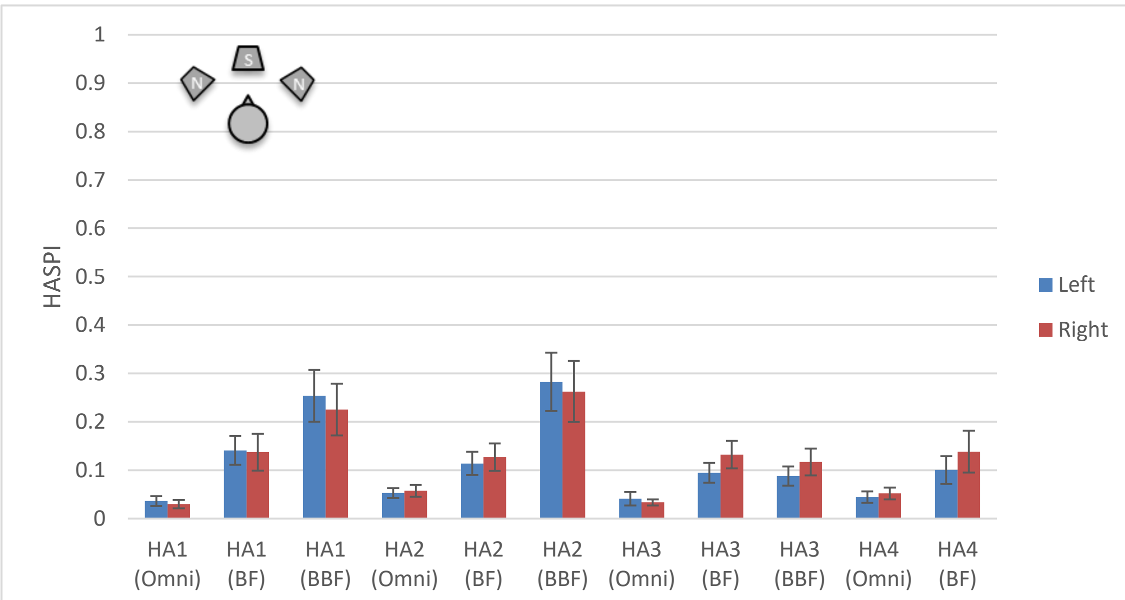


Figure 4.6: HASPI of speech in pink noise at 0 dB SNR, with speech originating from 0° and the noise from 45° and 315° recorded in the sound booth.

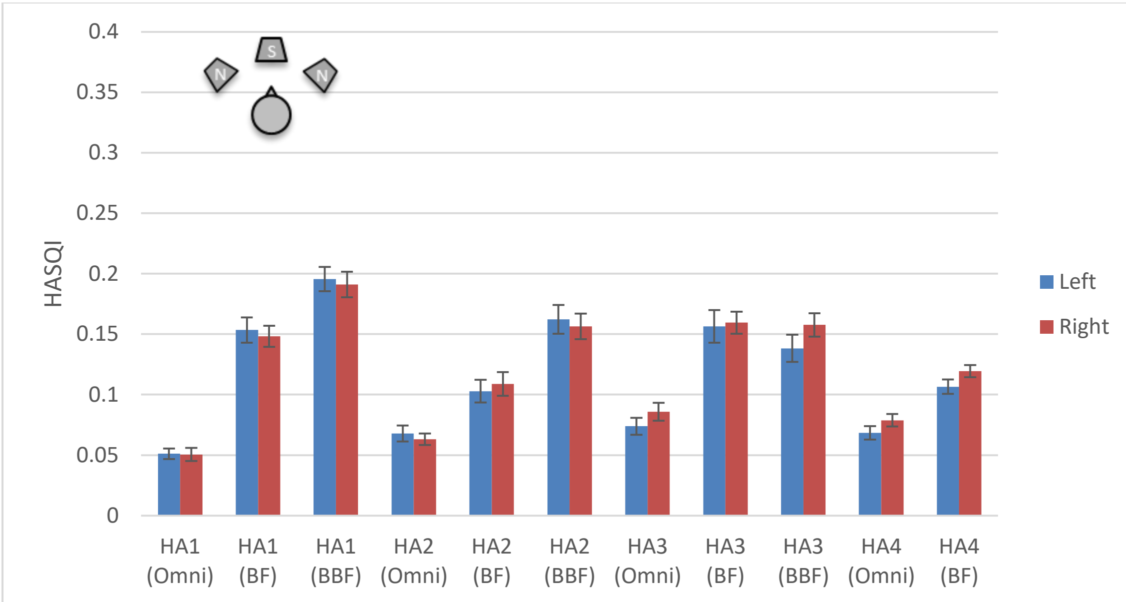


Figure 4.7: HASQI of speech in pink noise at 0 dB SNR, with speech originating from 0° and the noise from 45° and 315° recorded in the sound booth.

The HASPI and HASQI scores of pink noise, this time with the noise source at 90° and 180° as seen in Figure 4.8, can be interpreted differently depending on the weighting of the left and right scores. Hearing Aid 2, for example, sees a decrease in HASPI in the left ear and an increase in HASPI in the right ear on switching to a binaural beamformer. Hearing Aid 3 sees the opposite effect, where the left and right ear scores grow more extreme on switching to the binaural beamformer. This is the result of a difference in processing strategy between these two hearing aids, however given the difference in the left and right score it is difficult to effectively evaluate them and say definitively which one is more effective in improving predicted speech quality. Combining the left and right scores for each hearing aid in a perceptually relevant way, one of the goals of the thesis, would go a long way in aiding this comparison.

Regardless, both changes in predicted performance are fairly small. As this is a case with more localized noise than the previous two conditions, a similar conclusion to the first noise condition is likely where as long as the speech and noise sources have significant spatial separation, the binaural beamformer loses its advantage over the monaural beamformer.

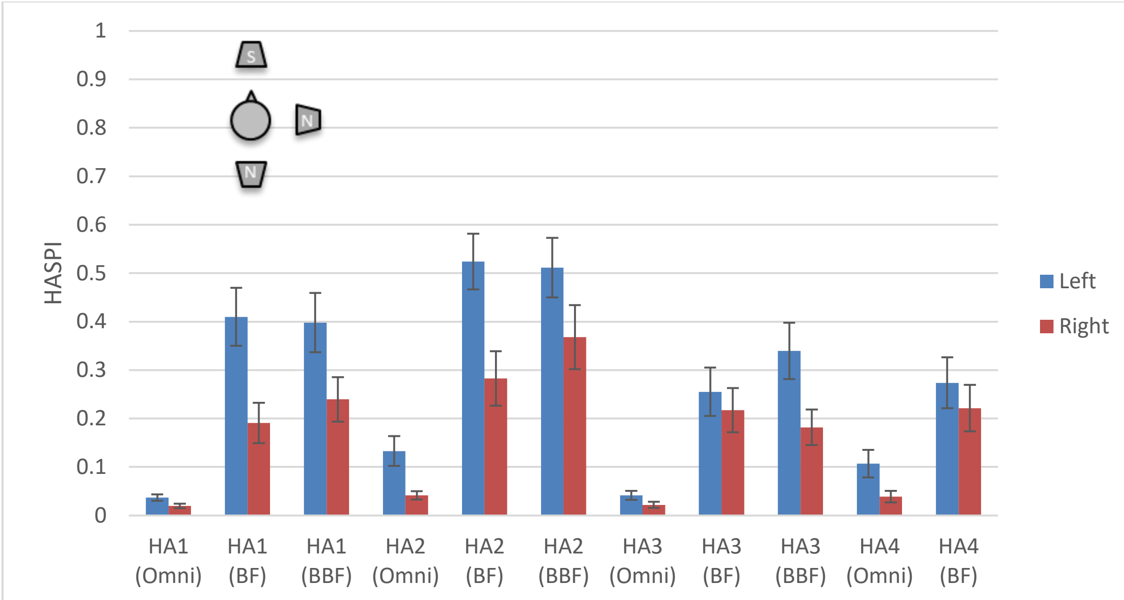


Figure 4.8: HASPI of speech in pink noise at 0 dB SNR, with speech originating from 0° and the noise from 90° and 180° recorded in the sound booth.

The final directionality condition has noise originating from 90° and 270°, similar to the first condition, however the speech source is now located at a 45° angle from the front-facing HATS.

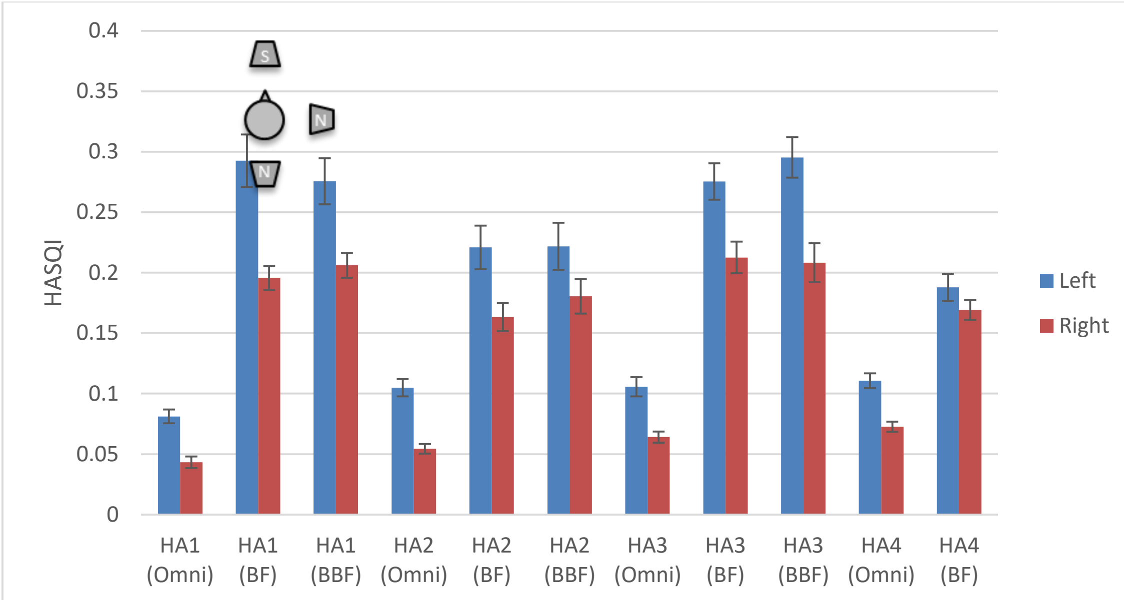


Figure 4.9: HASQI of speech in pink noise at 0 dB SNR, with speech originating from 0° and the noise from 90° and 180° recorded in the sound booth.

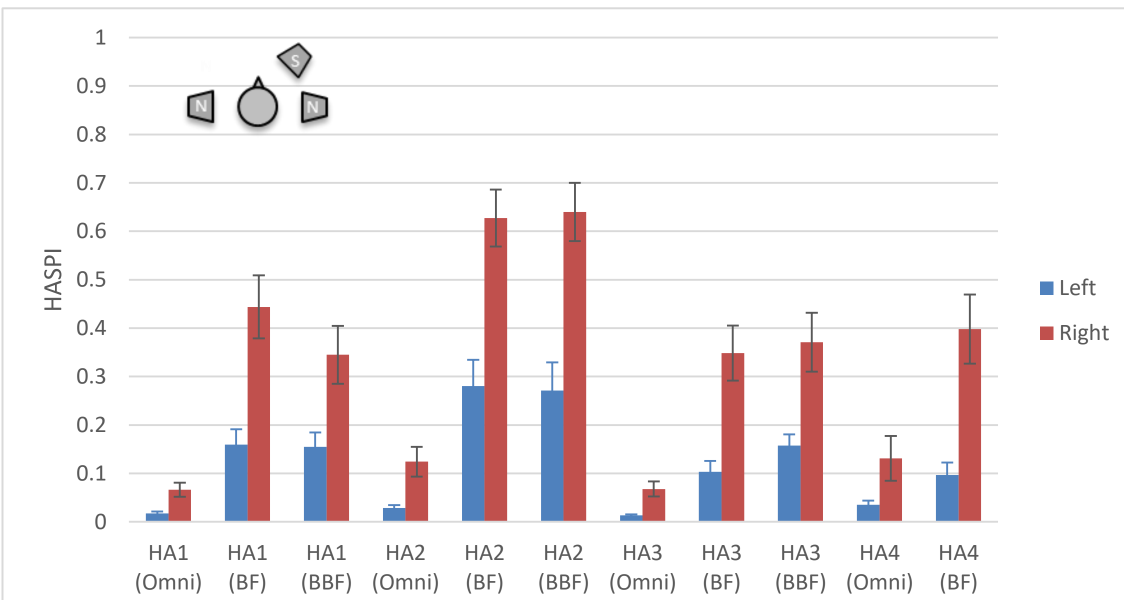


Figure 4.10: HASPI of speech in pink noise at 0 dB SNR, with speech originating from 45° and the noise from 90° and 270° recorded in the sound booth.

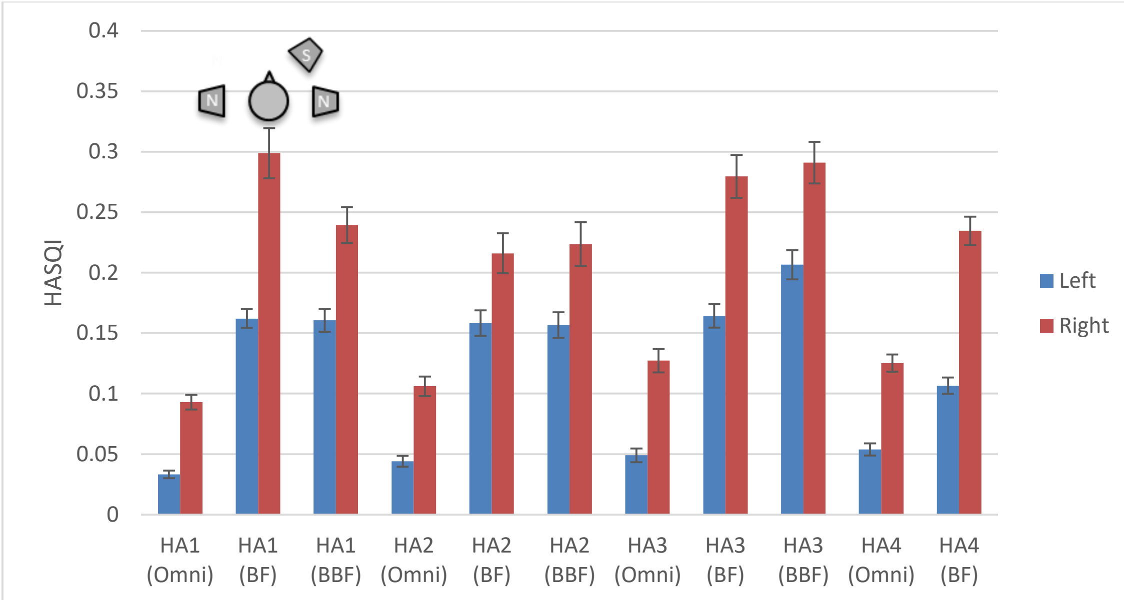


Figure 4.11: HASQI of speech in pink noise at 0 dB SNR, with speech originating from 45° and the noise from 90° and 270° recorded in the sound booth.

The discrepancy in the left and right scores due to the rightward angle of the speech source are made clear in Figure 4.10 and Figure 4.11. Angling the speech source 45° towards one ear has a clear benefit to the predicted quality and intelligibility of that ear, while negatively effecting the other. However by moving the speech source to a 45° angle, the predicted quality and intelligibility benefits of a binaural beamformer are lost in all three hearing aids with binaural beamforming programs. Notably, Hearing Aid 1 not only has no predicted quality or intelligibility benefit to the binaural beamformer when speech originates from a 45° angle, but there is actually a drop in both scores, likely due to the narrowing of a beamformer which may hinder the predicted quality and intelligibility scores in cases where the speech source is now outside the narrowed beamformer. Additionally, while there is a large discrepancy between the left and right ears of Hearing Aid 2 in this condition, the right ear has the best HASPI score of all conditions. Once again, the need for a comprehensive weighting function for both ears is demonstrated.

Switching the noise type from pink noise to cafeteria noise, an ambience of multi-talker babble meant to simulate a busy restaurant or café, or the SNR from 0 dB to 5 dB, did not significantly change the pattern of the HASQI or HASPI scores between brands and programs beyond a flat increase or decrease depending on the condition. The average HASQI and HASPI can be seen in Figure 4.12 and Figure 4.14, where the left and right scores of every monaural beamforming program were averaged together for each noise, SNR and directionality condition in the sound booth to view how the different conditions affected each hearing aid program on the whole. The equivalent figures for the binaural beamforming programs can be seen in Figure 4.13 and Figure 4.15.

Generally, higher SNR corresponded with higher HASQI and HASPI scores, which follows logically since a lower noise level will lead to less distortion to the original speech signal. Likewise, generally pink noise either corresponded with higher HASQI and HASPI scores compared to cafeteria noise or else there was no discernible difference. As pink noise is not statistically time-variant, it is possible to filter it out using statistical noise reduction methods as opposed to solely directional methods leading to stronger HASQI and HASPI scores than cafeteria noise in conditions where directional noise reduction is not possible. In areas where this is not the case, the averages HASQI and HASPI scores for speech in pink noise versus cafeteria noise are still within a standard error.

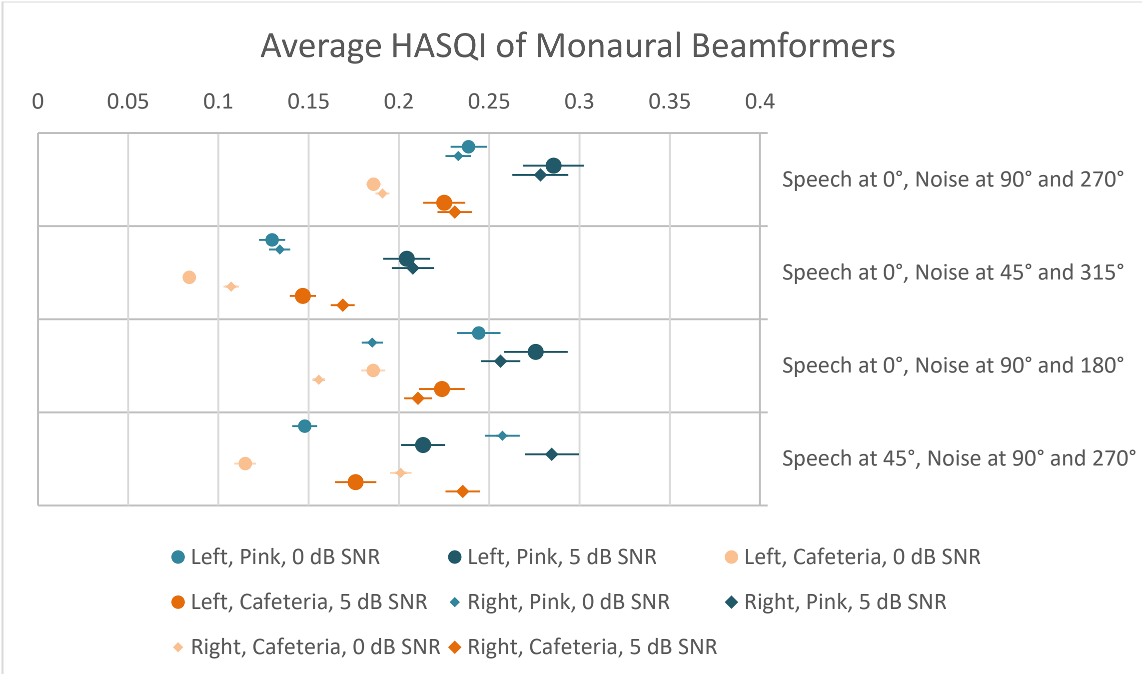


Figure 4.12: Average HASQI scores of monaural beamforming programs across all brands.

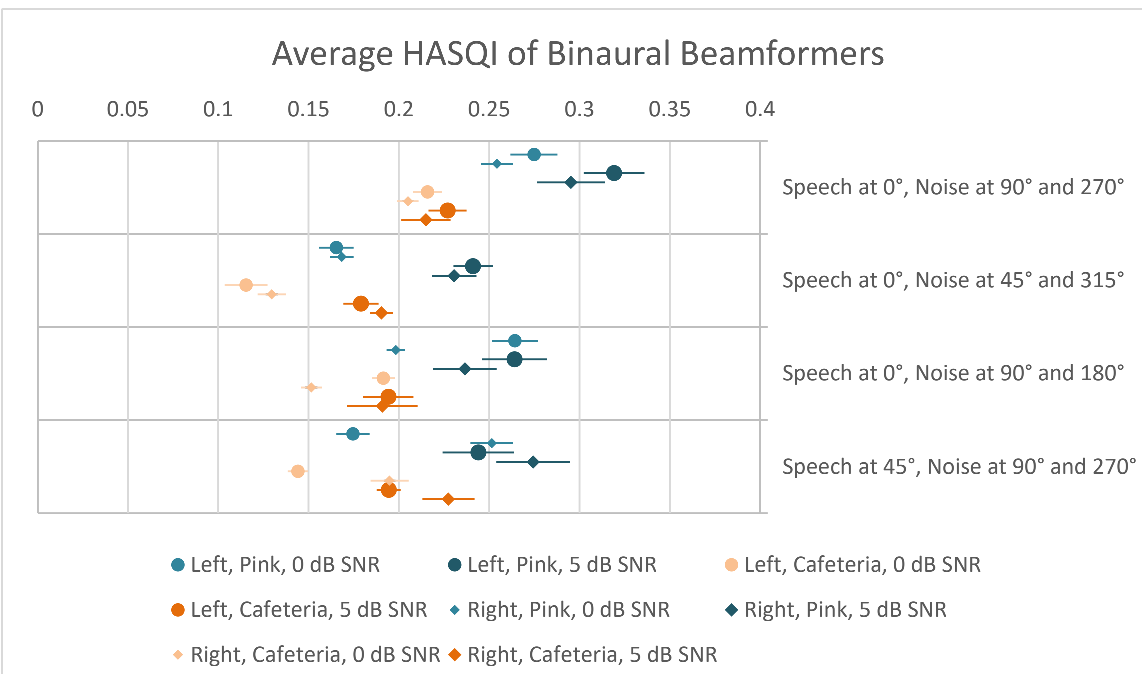


Figure 4.13: Average HASQI scores of binaural beamforming programs across all brands.

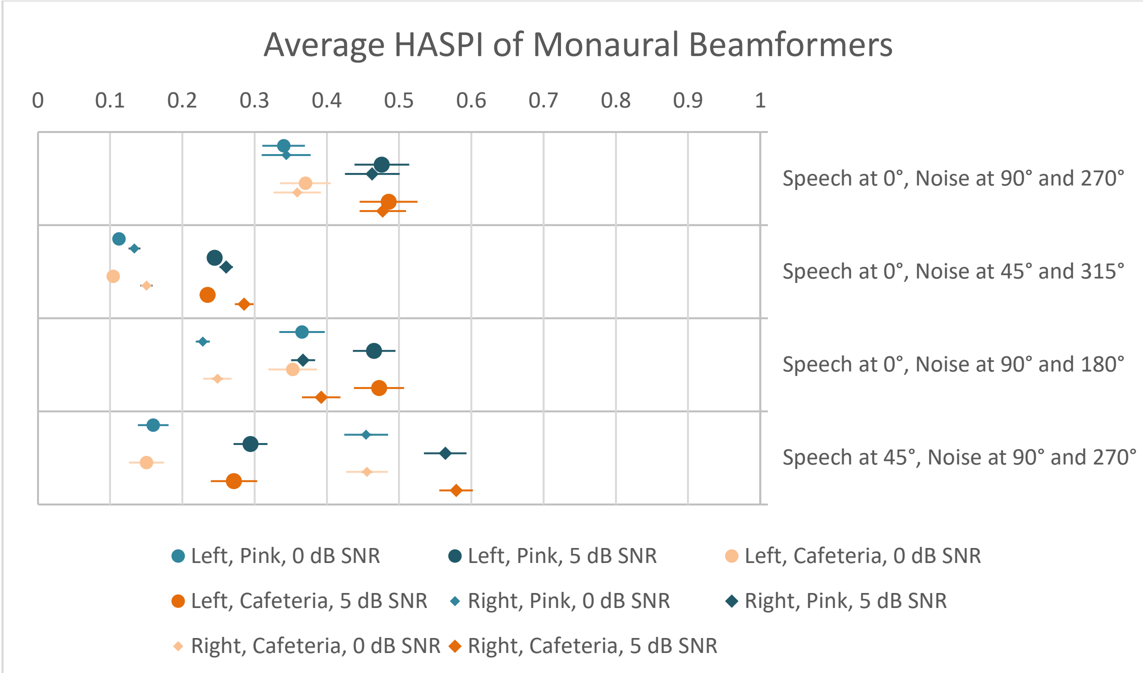


Figure 4.14: Average HASPI scores of monaural beamforming programs across all brands.

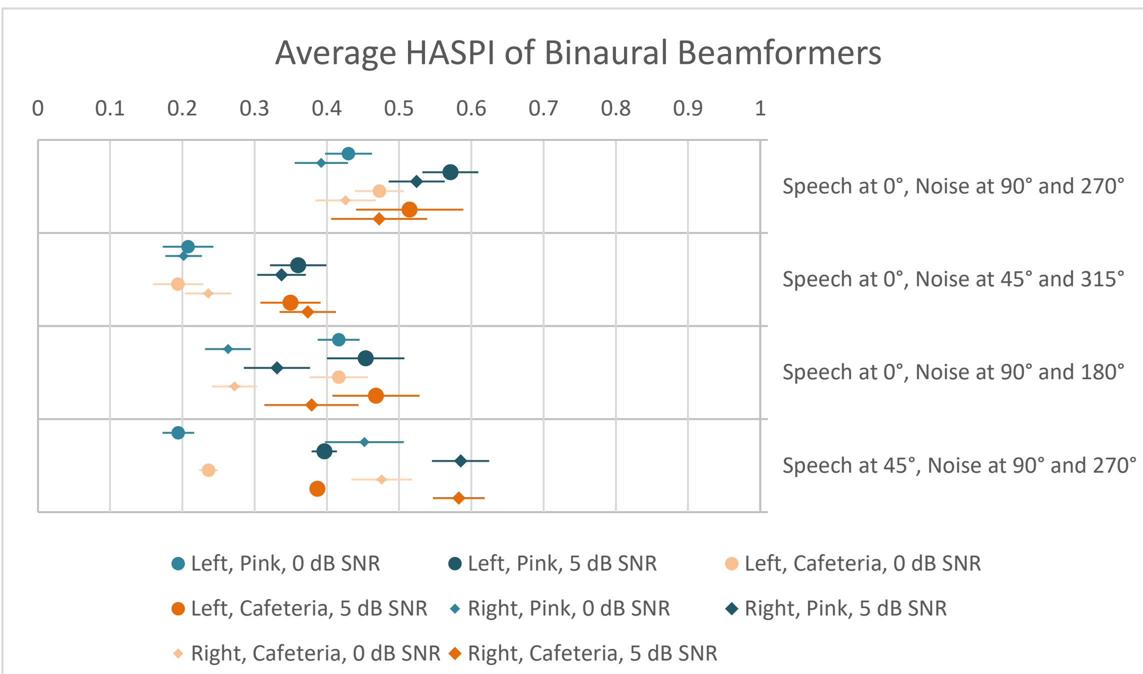


Figure 4.15: Average HASPI scores of binaural beamforming programs across all brands.

4.2.2 Reverberation Chamber

As the performance of the predicted speech quality and intelligibility scores were dependent on the spatial locations of the speech source and noise source, testing the hearing aids in reverberant conditions which negatively effects sound localization was key. Previous studies have shown the performance of binaural beamformers in speech recognition to be highly dependent on reverberation [33], [34]. The two conditions with the most notable differences between the measurements taken in the sound booth and reverb chamber was when the speech source originated from a 0° angle and the noise sources originated from a 45° and 315° angle, and when the speech source originated from a 45° angle and the noise source originated from a 90° and 270° angle. Respectively, these are the conditions where the binaural beamforming programs showed the greatest improvement over the monaural beamforming programs, due to the close proximity of the noise and speech sources, and the worst improvement, due to the decentering of the speech source out of the hearing aid's narrowed beamformer.

Comparing Figure 4.17, which displays a baseline condition in the reverb chamber similar to Figure 4.5 where the speech source originates from 0° and the noise sources originate from 90° and 270° with Figure 4.5, there is a drop in all HASQI scores in measurements taken in the reverb chamber versus the sound booth, however similar patterns emerge with a jump in predicted quality and intelligibility with the introduction of a monaural beamforming program, and similarly a minor improvement with the introduction a binaural beamforming program, this time in Hearing Aids 1 and 2, likely attributable to stronger de-reverberation processing in the noise reduction algorithm of these hearing aids. Also notable in Figure 4.17 is the performance of Hearing Aid 4, which remained competitive with the other hearing aid programs despite its poor HASQI score in the sound booth.

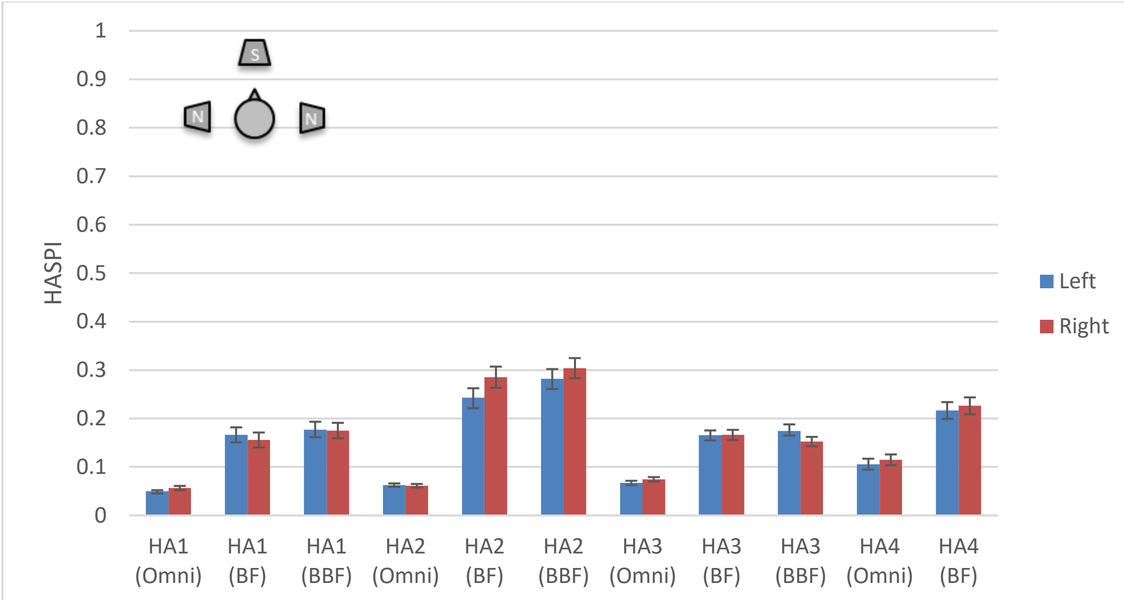


Figure 4.16: HASPI of speech in pink noise at 0 dB SNR, with speech originating from 0° and the noise from 90° and 270° recorded in the reverb chamber.

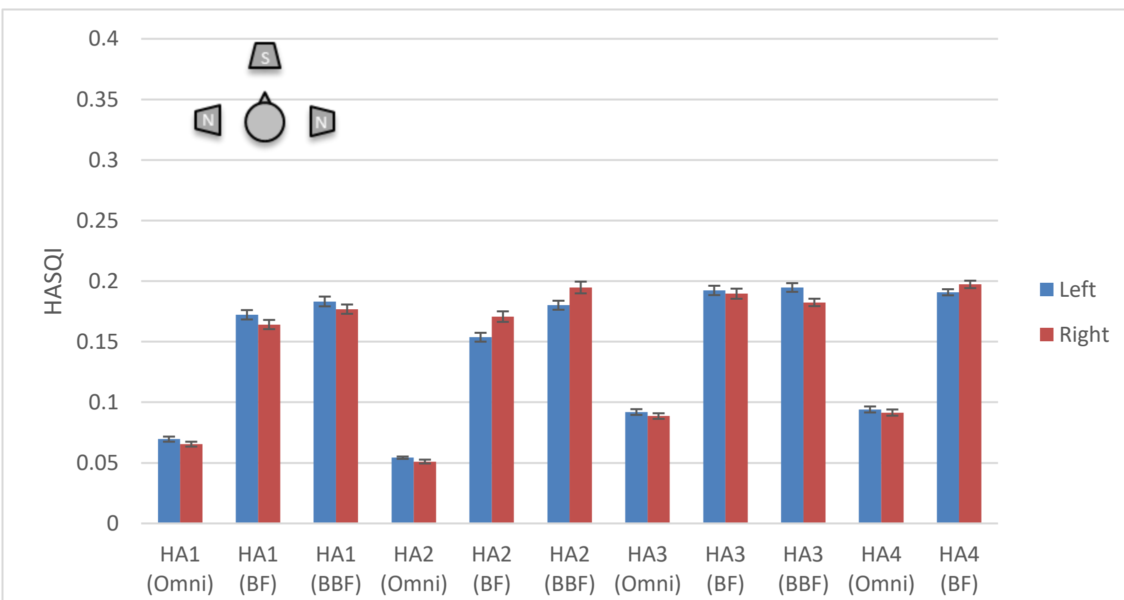


Figure 4.17: HASQI of speech in pink noise at 0 dB SNR, with speech originating from 0° and the noise from 90° and 270° recorded in the reverb chamber.

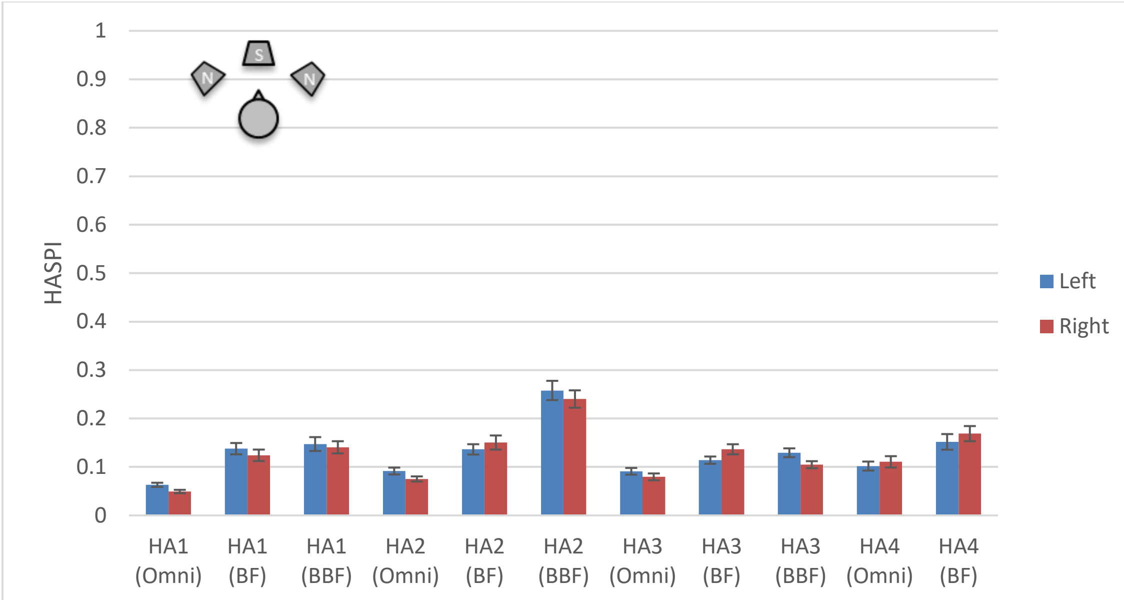


Figure 4.18: HASPI of speech in pink noise at 0 dB SNR, with speech originating from 0° and the noise from 45° and 315° recorded in the reverb chamber.

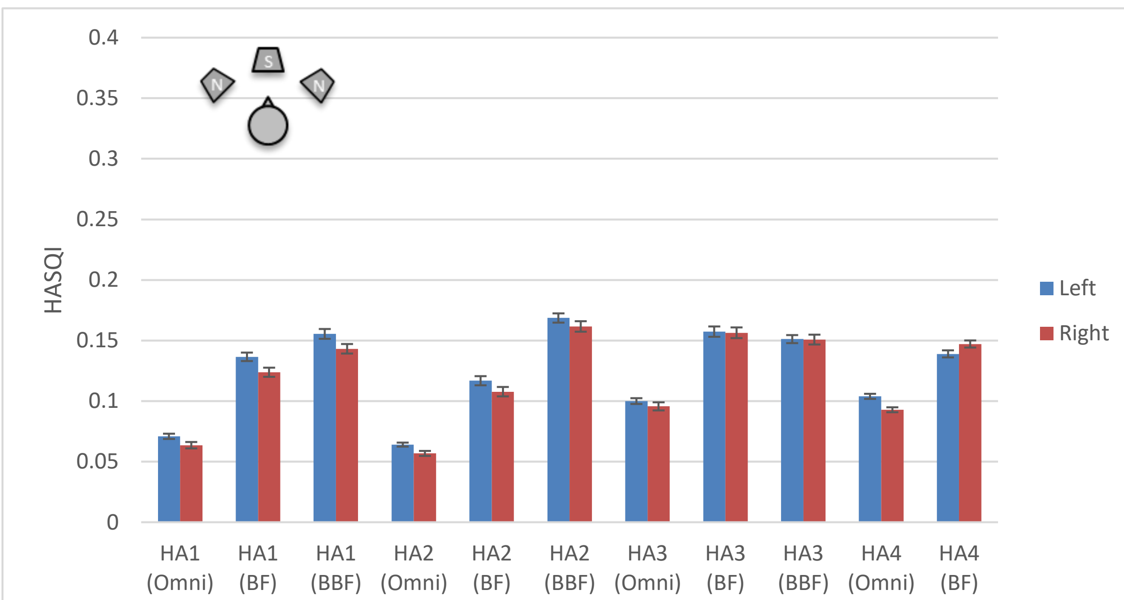


Figure 4.19: HASQI of speech in pink noise at 0 dB SNR, with speech originating from 0° and the noise from 45° and 315° recorded in the reverb chamber.

By comparing Figure 4.16 and Figure 4.17 with Figure 4.19 and Figure 4.19 it is possible to see a greater improvement with the binaural beamforming programs when the proximity of the noise sources is closer to the speech sources, particularly with Hearing Aid 2. This was a similar pattern seen with the measurements taken in the sound booth. Once again, placing the noise sources in closer proximity to the speech source negatively affected the predicted speech quality and intelligibility scores more than the binaural beamformers could make up for, however they still improved the performance significantly beyond what was capable with the monaural beamformer in the same condition.

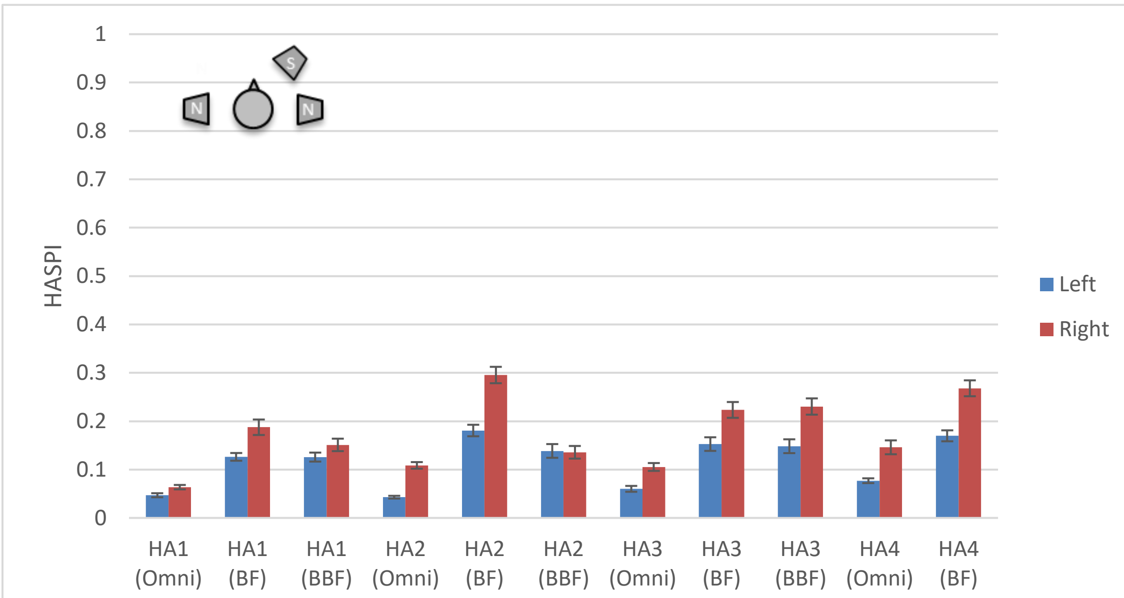


Figure 4.20: HASPI of speech in pink noise at 0 dB SNR, with speech originating from 45° and the noise from 90° and 270° recorded in the reverberation chamber.

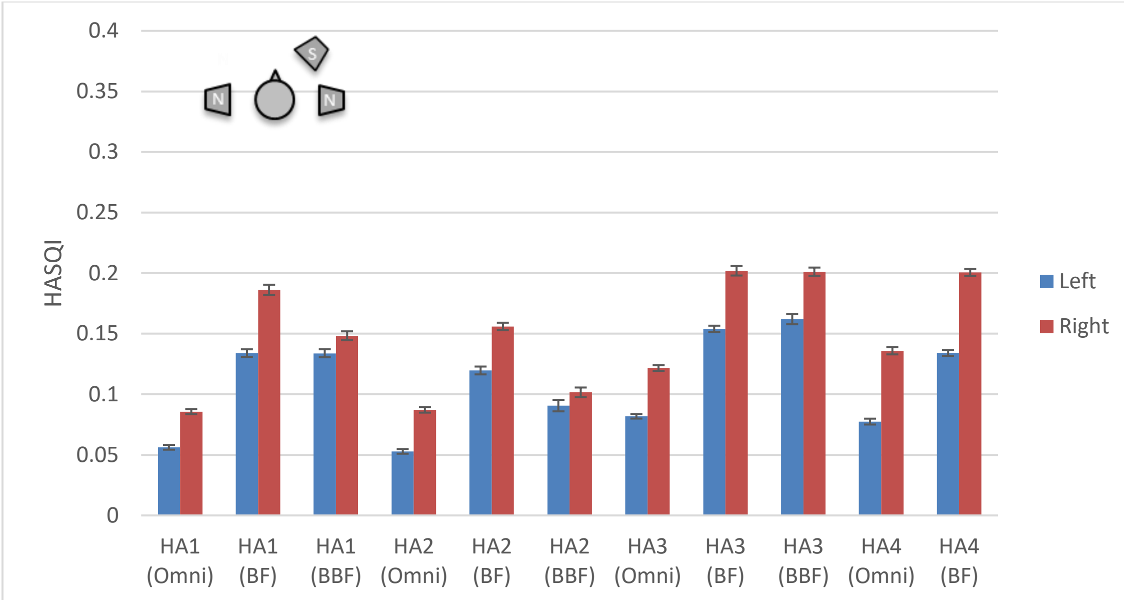


Figure 4.21: HASQI of speech in pink noise at 0 dB SNR, with speech originating from 45° and the noise from 90° and 270° recorded in the reverberation chamber.

By a similar token, in Figure 4.20 and Figure 4.21 it is possible to see how the binaural beamforming programs perform in predicted speech quality and intelligibility in conditions unsuited for their capabilities. With the speech source at a 45° angle, Hearing Aid 3's binaural beamformer performed similar to its monaural beamformer whereas Hearing Aids 1 and 2 both showed a significant decrease in predicted quality and intelligibility. Once again, this is likely due to the narrowed beamformer of the binaural program, which only amplifies signals within a certain range of angles. Once the speech source leaves that narrow range, it is no longer amplified. Hearing Aid 2 did not initially show a drop in predicted speech quality or intelligibility when measured in a sound booth, but in the reverberation chamber it did. This is likely due to the reverberation of the reverberation chamber negatively affecting Hearing Aid 2's ability to localize the speech and therefore not correcting its beamformer in any way.

Hearing Aid 1 also shows an interesting pattern in this condition where with the introduction of the binaural beamformer, the left ear HASQI score improved where the right ear HASQI score decreased, decreasing the disparity between the two ears scores. Since binaural beamformers use the weighted sum of the monaurally processed inputs from each device, it follows that some level of "equalization" would occur between them where the worse performing device would improve and the better performing device would worsen. In most cases, this is an attempt from the manufacturer to use the best of the two hearing aid signals. However as the weighted sum must use at least some component of the input signal from both devices in order to preserve localization cues, the better performing ear may experience a drop in performance as it is summed with the device experiencing more noise.

4.3 B&K HATS and CARL Comparison

As there was a significant decrease in performance in Hearing Aid 2 when the speech source was at a 45° direction, it was important to verify that the decrease was due to the narrow beam of the binaural beamformer and not due to the interference from either the electrical components or material of the HATS. A comparison experiment was therefore performed using a Canadian Audiology simulator for Research and Learning (CARL) to ensure similar results were gathered between the microphone-equipped HATS and the more anatomically accurate, and hollow CARL.

The CARL was fitted with the Real Ear Measurement system on the Audioscan Verifit 2 hearing aid measurement system, which uses probe tubes inserted into the ear canal to take hearing aid recordings. The CARL was then affixed with Hearing Aid 2 fitted to an N4 audiogram in the reverb chamber and recorded in an omnidirectional program (Omni), the monaural beamforming program (BF), and the binaural beamforming program (BBF) with speech-shaped noise coming from 90° and 270° and speech coming from 0° and then again with speech at 45°.

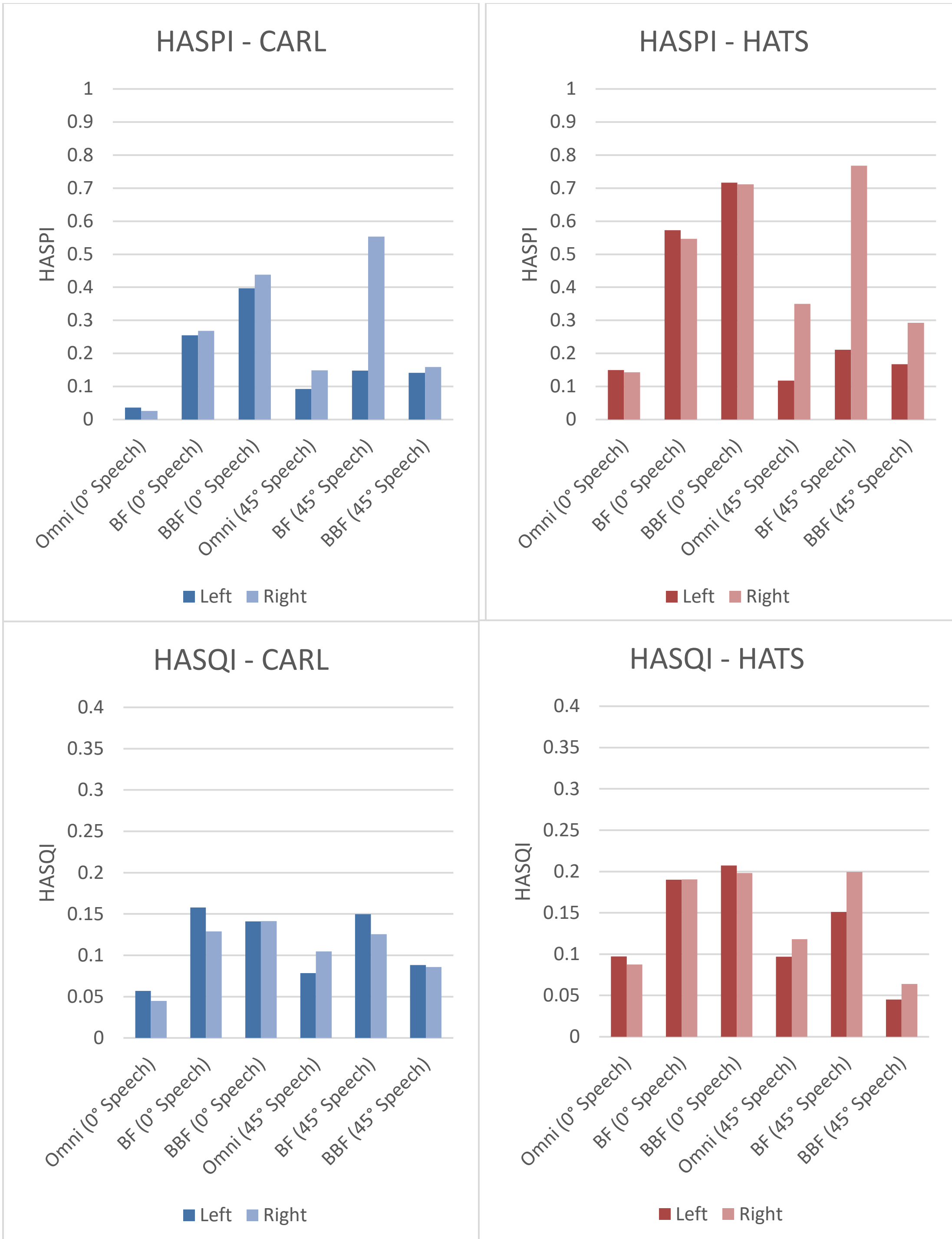


Figure 4.22: Comparison of Hearing Aid 2 HASPI and HASQI for HATS and CARL.

As the Verifit 2 was not calibrated to the same level as the HATS, the absolute value of HASQI and HASPI could not be compared. However the CARL recordings still showed the same decrease in performance when speech was at a 45°, as seen in Figure 4.22, consistent with the same measurement done with the HATS. Therefore the decrease could not be attributed to material or electrical interference from the HATS.

4.4 Summary

Electroacoustic measurements of hearing aids allowed for several insights into the performance of binaural beamformers. First, the predicted speech quality and intelligibility of the beamformers was dependent on the direction of the speech and noise source. When the noise sources were at 90° and 270°, the binaural beamformer provided a small predicted quality and intelligibility improvement over the monaural beamformer. The benefit over the monaural beamformer increased when the noise sources were moved to 45° and 315°. When the noise sources were more localized at 90° and 180°, the predicted quality and intelligibility benefit was negligible. Finally, when the speech source was rotated to a 45° angle, the binaural beamformer provided no benefit over the monaural beamformer and in some cases decreased the HASPI and HASQI score.

The performance of the binaural beamformers was also dependent on the reverberation of the environment. Particularly, the predicted speech quality and intelligibility drop when the speech source was at 45° increased. Changing the noise type from pink to cafeteria noise or the SNR from 0 to 5 dB did not change the results significantly.

Chapter 5

5 Behavioural Analysis

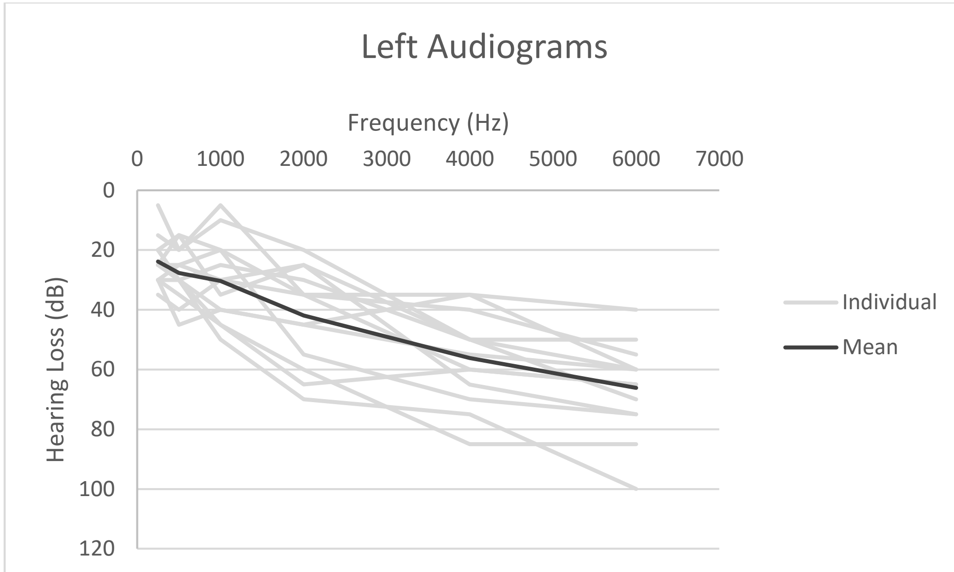
To gain a thorough understanding of the effect of binaural beamforming algorithms on speech quality as well as to validate the objective speech quality metric HASQI with binaural beamformers, subjective ratings of speech processed by binaural beamforming hearing aids were collected. Thirteen hearing impaired participants were recruited to take part in the study. Similar to the electroacoustic measurements, recordings of the hearing aids were first made according to the conditions detailed in Table 5.1 programmed to each participant's audiograms. Recordings were then presented to the participants using the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) methodology and ranked from zero to one hundred.

5.1 Methods

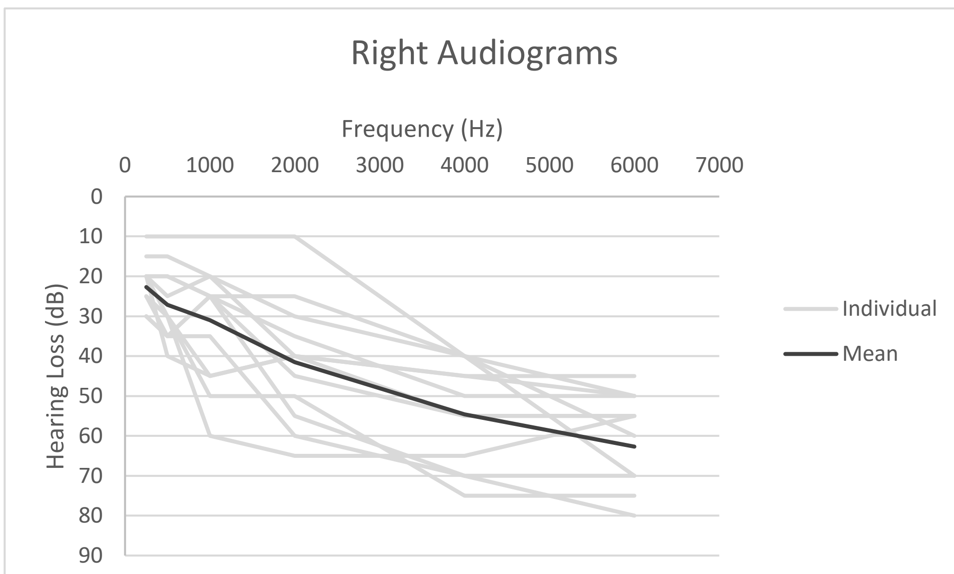
5.1.1 Participants

Participants were recruited through a pre-existing database of hearing impaired listeners who frequently participate in studies through the NCA. A total of 13 participants were brought in for the study with mild to moderate hearing loss based on the Pure Tone Average (PTA); all were experienced hearing aid users and ranged in age between 60 to 86 years with a mean age of 73. The ages of the participants are listed in Appendix A, while the individual and average audiograms of all participants are shown in Figure 5.1. The study was approved the Western University Health Sciences Research Ethics Board (HSREB), which can be viewed in Appendix D.

Hearing aids were then programmed according to the given participant's audiogram and then verified with an Audioscan Verifit 2 based on the participant's Real Ear to Coupler Difference (RECD) values. Each of the participant's audiogram and RECD values were adjusted based on the HATS difference displayed in Appendix B.



a)



b)

Figure 5.1: Individual and average audiogram of participants for the a) left ear and b) right ear.

5.1.2 Speech and Noise Conditions

With the hearing aids programmed to each participant's audiograms, recordings were made in each of the conditions detailed in Table 5.1.

Table 5.1: Hearing aid recording conditions for behavioural measurements.

Speech and Noise Directions	**Hearing Aids**	**Programs**	**Noise Types**	**SNR**	**Rooms**
• Speech at 0°, Noise and 90°, 180° and 270° • Speech at 0°, Noise at 90° and 180° • Speech at 45°, Noise at 90°, 180° and 270° • Speech at 0°, No Noise*	• Hearing Aid 1 • Hearing Aid 2 • Hearing Aid 3 • Hearing Aid 4	• Omnidirectional • Monaural Beamformer • Binaural Beamformer† • Better Ear†	• Speech-shaped • Cafeteria	• 0 dB	• Sound booth • Reverb Chamber

*In conditions with no noise, the Noise Types condition was not necessary

†Binaural beamforming programs were not available on Hearing Aid 4, and so a Better Ear program was used instead. Better Ear takes the signal with the highest SNR and outputs it to both ears.

A few changes were made between the conditions of the electroacoustic recordings and the recordings made for the behavioural tests. First, it was known that some number of conditions would have to be removed for logistical reasons. As the tests were to be done with participants, each additional condition added exponentially more recordings each participant would have to listen to. It was decided it would be unreasonable to ask for participants to listen to any more than 200 recordings in one sitting, therefore with four hearing aid brands and three programs each, the number of total listening conditions was reduced to 16. The reasoning behind each condition change or removal is detailed below.

Pink Noise to Speech-Shaped Noise: The pink noise condition was changed to speech-shaped noise. While it was important for one of the noise conditions to be spectrally stable in order to contrast with the cafeteria noise, speech-shaped noise was chosen to better mask the spectral features of the underlying speech signal.

Removal of Noise from 45° and 315° Condition: The second noise direction condition was removed by the process of elimination. Noise from 90° and 270° needed to be kept as a baseline, and speech from a 45° degree angle was kept to see the effect of an off-angle speech source on the quality. Between the noise at 90° and 180° condition and the noise at 45° and 315° condition, the former was kept as it had a greater discrepancy between the left and right HASQI scores and therefore would provide stronger data for a weighting function between the two ears.

Removal of 5 dB SNR Condition: The 5 dB SNR condition was removed to ensure there was enough room to test multiple noise types.

Addition of Noise at 180° in Noise Direction 1 and 3: A third noise source at 180° was added to noise directions 1 and 3 to increase the dispersion of sound outside the speech source.

Addition of Better Ear Program for Hearing Aid 4: As Hearing Aid 4 did not have a binaural beamformer, a third Better Ear program was added to see how it compared to the binaural beamformers.

Addition of Speech in No Noise Condition: A speech in quiet condition was added both as a reference for MUSHRA and for statistical reliability analysis.

5.1.3 MUSHRA Test

Once recordings of the four hearing aids programmed to each participant’s audiogram were made in each of the listed Table 5.1 conditions, participants were brought to the NCA and instructed to rank each of the recordings according to the MUSHRA methodology. As seen in Figure 5.2, MUSHRA utilizes a single screen where each letter corresponds to a different hearing aid program, while each screen corresponds to a specific noise, room, and directionality condition. By clicking on a letter, the participant can listen to the corresponding recording, which are randomly ordered on the screen, and then rank it using the slider from zero to 100. Participants were also free to adjust the volume to a comfortable level using the slider at the top of the screen. Participants ranked 12 hearing aid programs per screen for 16 screens, where two screens were speech recordings made in identical, no noise conditions for internal statistical analysis.

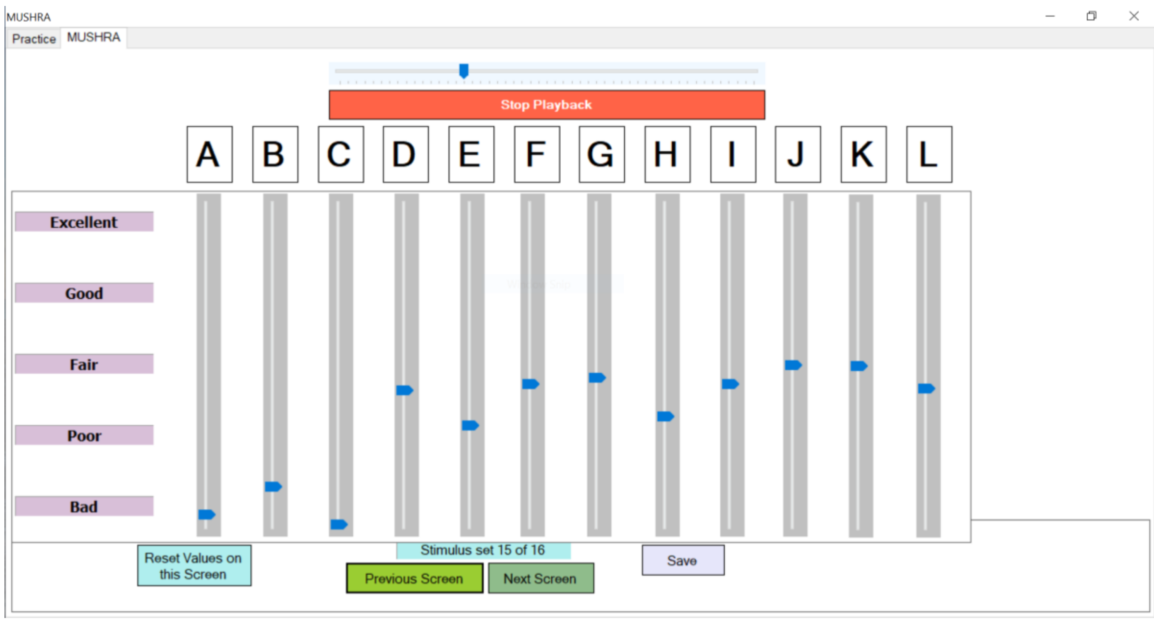


Figure 5.2: The MUSHRA program participants used to rank hearing aid recordings.

5.2 Behavioural Data Analysis Results

Subjective ratings were compiled alongside the HASQI score for the respective ranked recording and conditionally averaged. In other words, the subjective rating and HASQI score were averaged with all other participant's rating and HASQI score for the same noise, room, and directionality condition.

Intra-participant reliability was measured with the correlation coefficients between the two identical, no noise conditions for each participant and was found to have a range of -0.1672 to 1 for recordings made in the sound booth and -0.1664 to 0.9462 for recordings made in the reverberation chamber. A large range of correlation coefficients can be ascribed to the lack of variation between the two no noise conditions. Inter-participant reliability was then measured with Cronbach's alpha, which was found to be 0.9246 for recordings made in the sound booth and 0.9230 for recordings made in the reverberation chamber. When the data was restricted to not include recordings made in the no noise condition which were consistently ranked highly, Cronbach's alpha was calculated to be 0.8199 for recordings made in the sound booth and 0.6871 for recordings made in the reverberation chamber.

As seen in Figure 5.3, the left-right mean HASQI scores were correlated with the corresponding subjective ratings, providing an R squared value of 0.7504. The R squared value rose even higher when scores were restricted to sound booth recordings with little to no reverb, rising to 0.9017 as seen in Figure 5.4. With the reverb chamber recordings alone, the R squared value fell to 0.7327 as seen in Figure 5.5. Overall, the left-right mean HASQI score provided a good indicator of subjective ratings when the reverberation of the environment is controlled.

A gap in data points between subjective ratings of 62 and 75 is visible in the Figure 5.3, Figure 5.4, and Figure 5.5, which illustrates the predicted quality gap between the speech in no noise conditions versus the speech in 0 dB SNR conditions. The grouping of data points at the top in Figure 5.3 corresponds with recordings made in the sound booth, while the grouping of data points closer to the bottom corresponds with recordings made in the reverberation chamber, highlighting the impact of reverberation on perceived speech quality even in the absence of any background noise.

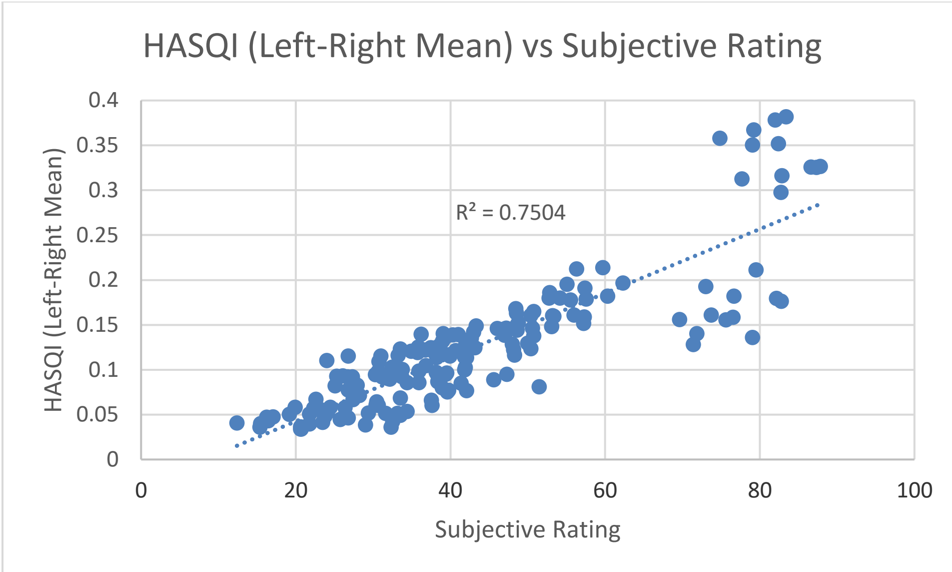


Figure 5.3: Correlation of the conditional averages of the mean HASQI score between left and right ears and the corresponding recording's subjective ranking.

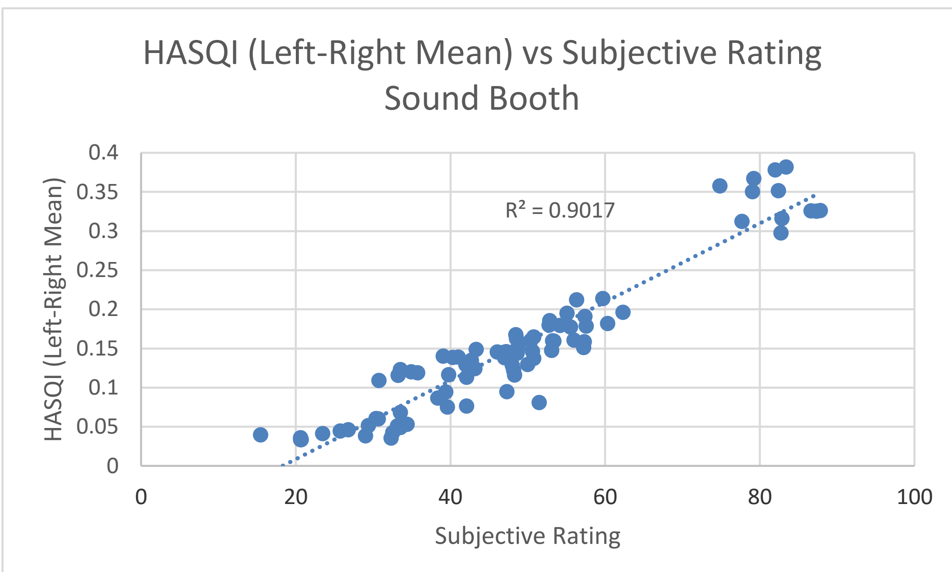


Figure 5.4: Correlation of the conditional averages of the mean HASQI score between left and right ears and the corresponding recording's subjective ranking for recordings made in sound booth.

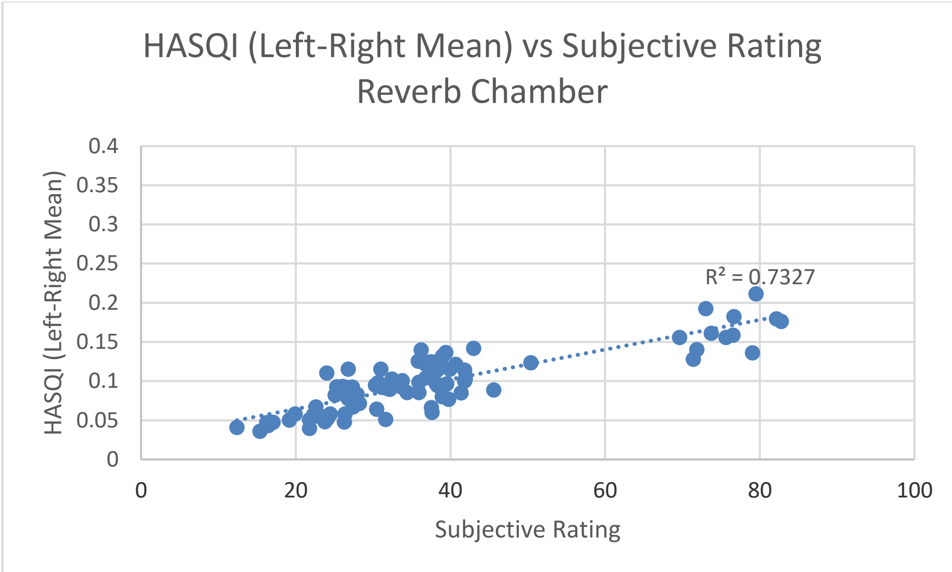


Figure 5.5: Correlation of the conditional averages of the mean HASQI score between left and right ears and the corresponding recording's subjective ranking for recordings made in reverb chamber.

5.2.1 Sound Booth

Comparing the conditionally averaged results of the subjective ratings with the conditionally averaged HASQI scores of the corresponding recordings allows for new information regarding the effect of binaural beamforming on speech quality. Figure 5.6 shows the conditionally averaged subjective rating, the conditionally averaged left-right maximum HASQI values, and the conditionally averaged left-right mean HASQI values respectively when speech is presented at the HATS at 70 dB SPL from 0° while speech-shaped noise is presented from 90°, 180° and 270° at 0 dB SNR in the sound booth. As the noise is equally spaced around the HATS and statistically stable, this condition becomes the new baseline for comparison.

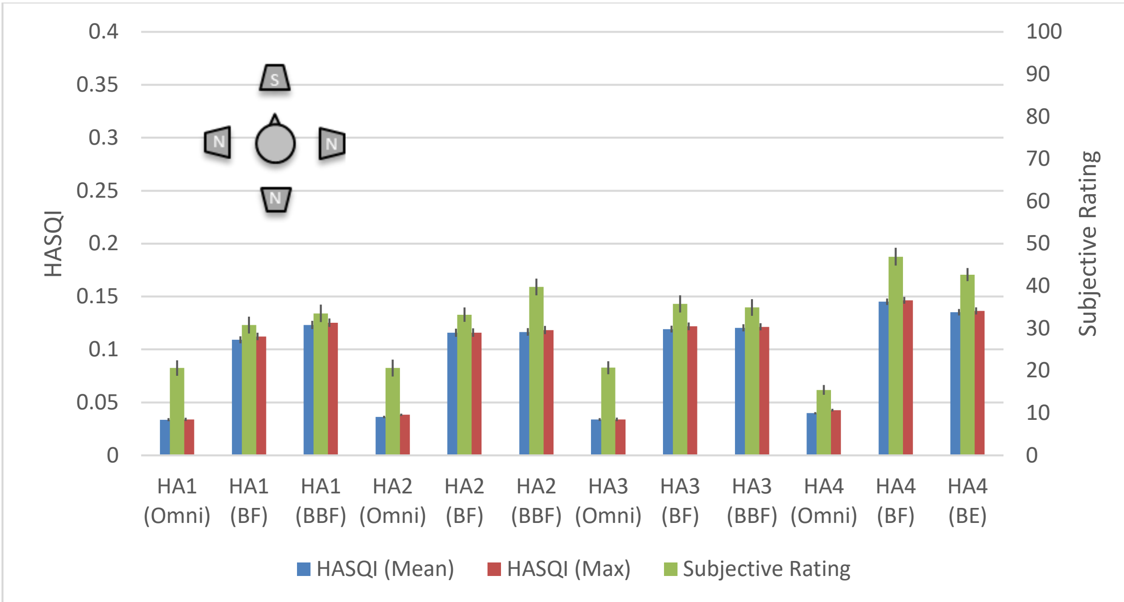


Figure 5.6: Conditionally averaged subjective rating, left-right maximum HASQI score and left-right mean HASQI score of speech in speech-shaped noise at 0 dB SNR, with speech originating from 0° and the noise from 90°, 180° and 270° in the sound booth.

Overall, intra-hearing aid score patterns between programs are maintained between the subjective ratings and maximum HASQI. Both Hearing Aids 1 and 2 increase in rated speech quality and predicted speech quality from the omnidirectional program to the monaural beamformer, and then again to the binaural beamformer. Hearing Aids 3 and 4, however, see an increase in rated speech quality and predicted speech quality from the omnidirectional program to the monaural beamformer, but then Hearing Aid 3 maintains a similar score in the binaural beamformer and Hearing Aid 4 drops in performance for the Better Ear program. Likely this was a result of switching from pink noise to speech-shaped noise, which may cause more interference with the noise reduction algorithms for Hearing Aid 3. It also follows logically that in noise-symmetric conditions, the better ear program for Hearing Aid 4 would not see an increase in rated speech quality or predicted speech quality since both ears would perceive the same SNR.

The left-right maximum HASQI provided a closer similarity between the subjective ratings and the HASQI score versus the left-right mean HASQI in the case of hearing aid

3, where the left-right mean HASQI score saw an increase in predicted speech quality from the monaural beamformer to the binaural beamformer, whereas the subjective speech quality ratings and the left-right maximum HASQI score saw a decrease.

The pattern of inter-hearing aid performance was largely maintained between the rated speech quality and predicted speech quality scores. The best subjectively rated program, Hearing Aid 4's monaural beamformer, was also the best scoring HASQI maximum.

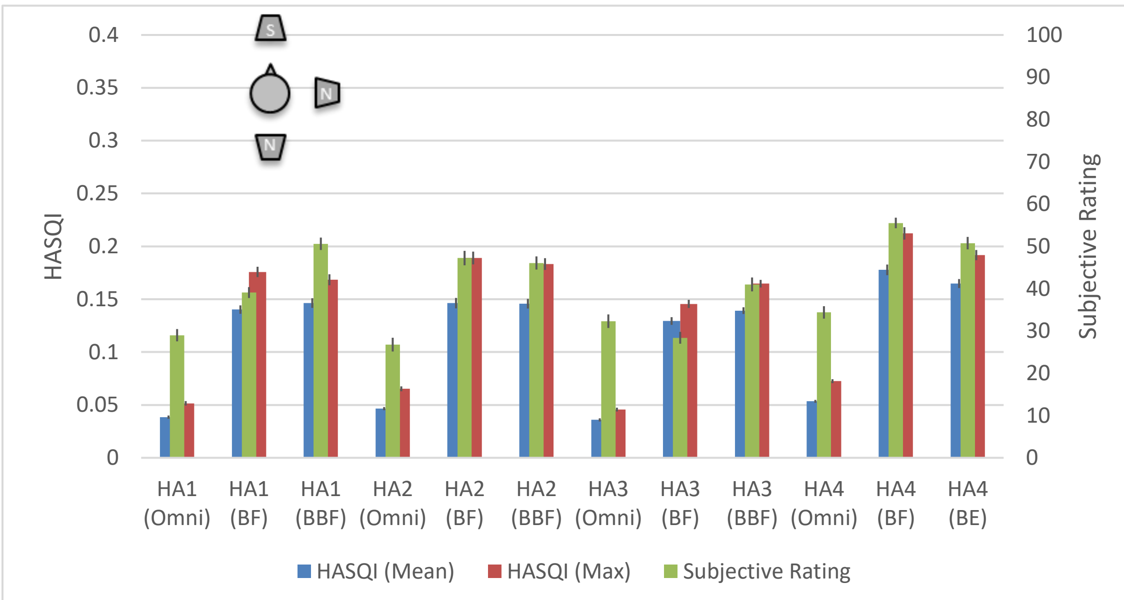


Figure 5.7: Conditionally averaged subjective rating, left-right maximum HASQI score and left-right mean HASQI score of speech in speech-shaped noise at 0 dB SNR, with speech originating from 0° and the noise from 90° and 180° in the sound booth.

The pattern of results between the subjective rating and HASQI scores in asymmetrical noise conditions can be viewed in Figure 5.7, which compares the results in when speech is presented at the HATS at 70 dB SPL from 0° while speech-shaped noise is presented from 90° and 180° at 0 dB SNR in the sound booth. The performance between metrics of certain hearing aids, such as Hearing Aid 2, maintained well between the HASQI scores and subjective ratings, with the monaural beamformer and binaural beamformer performing similarly and with much higher subjective rating and HASQI scores than the omnidirectional program. Hearing Aid 4 had similarly harmonious results, with the monaural beamformer performing best in both subjective rating and predicted speech quality and seeing a minor drop in both metrics when switched to the better ear program.

Hearing Aid 3 sees a discrepancy in the monaural beamformer, which drops even below the omnidirectional program in subjective ratings whereas the HASQI scores improve. As the omnidirectional program uses no noise reduction, and is used a baseline to compare the other programs, it is an outlier compared to the overall correlation of HASQI score to subjective ratings.

Hearing Aid 1 sees a discrepancy between the metrics when switched from the monaural beamforming program to the binaural beamforming program, as the left-right maximum HASQI scores see a drop on the switch. This contrasts with the subjective ratings which see an improvement on the switch from monaural beamforming to binaural beamforming. The left-right mean HASQI score follows the same pattern as the subjective ratings, with the binaural beamforming program again performing the best in predicted speech quality. This implies that the left-right mean HASQI score is a better indicator of subjective speech quality in conditions where the noise source is at 90° and 180°.

Additionally, while Hearing Aid 2's binaural beamforming program does not improve subjective ratings, no binaural beamforming program has an adverse effect on the subjective speech quality rating and most see an improvement.

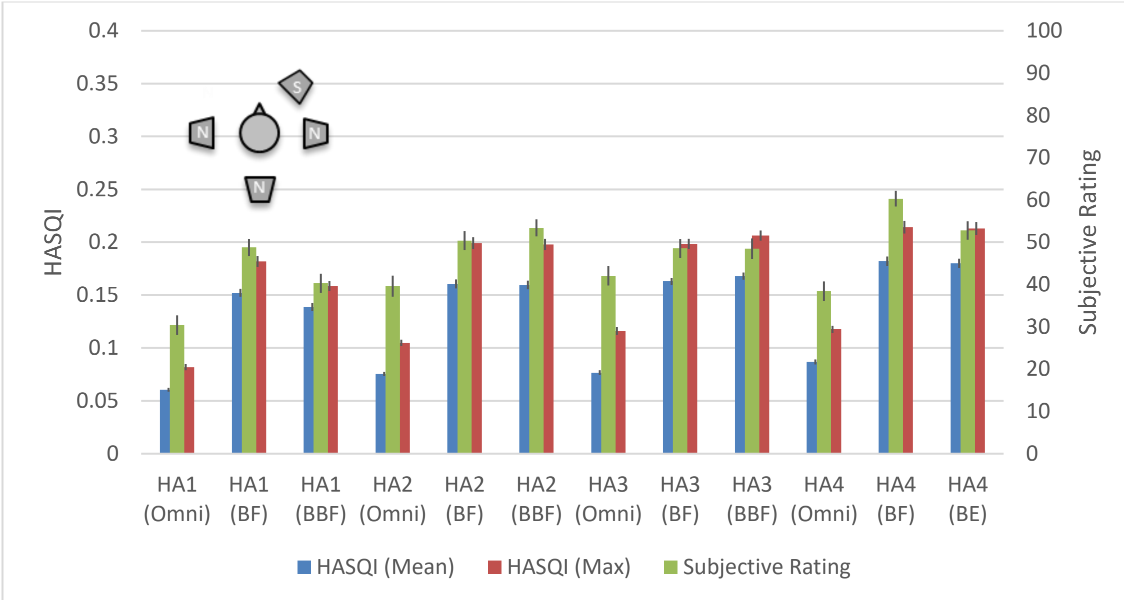


Figure 5.8: Conditionally averaged subjective rating, left-right maximum HASQI score and left-right mean HASQI score of speech in speech-shaped noise at 0 dB SNR, with speech originating from 45° and the noise from 90°, 180°, and 270° in the sound booth.

More information regarding the effect of asymmetry on binaural beamforming programs can be discerned from recordings made with the speech source at a 45° angle, as seen in Figure 5.8. The subjective ratings follow a similar pattern to the HASQI score, with Hearing Aids 2 and 3 seeing little to no benefit from the binaural beamforming program in either subjective rating or HASQI. Notably, Hearing Aid 1 sees a decrease in subjective rating for both subjective ratings and HASQI score. This aligns with the electroacoustic analysis in 4.2.1, where Hearing Aid 1 saw a decrease in HASQI score when switched to its binaural beamforming program when the speech source originated from a 45° angle.

Unlike the electroacoustic analysis in Section 4.2, switching from speech-shaped noise to cafeteria noise introduced significantly more variability into the results for recordings made in conditions where the speech source is at 45° and the noise source is at 90°, 180° and 270°, as seen in Figure 5.9. While the intra-hearing aid HASQI score patterns were largely maintained, with Hearing Aids 2, 3, and 4 seeing no significant change in predicted speech quality performance between the monaural and binaural programs and Hearing Aid 1 seeing a drop from the monaural beamformer to the binaural beamformer, the subjective ratings saw a drop from the monaural to binaural programs in hearing aids 1, 2 and 4. In Hearing Aid 1 specifically, the drop was significant enough that the binaural beamformer performed worse that the baseline omnidirectional program. While this seems unusual, given that the speech was out of the narrow range of the binaural beamformer it makes sense that the binaural beamformer would not perform as intended. Compounding this fact with the non-stochastic nature of the noise means that the hearing aids could not depend on a statistical analysis of the noise as a backup noise reduction technique, and therefore may have misinterpreted the cafeteria noise as a wanted signal. In other words, there was no discernible quality of the speech that separated in from the noise, either directionally or statistically. Still, the deviation of the subjective ratings from the HASQI scores make it an important component of the data set.

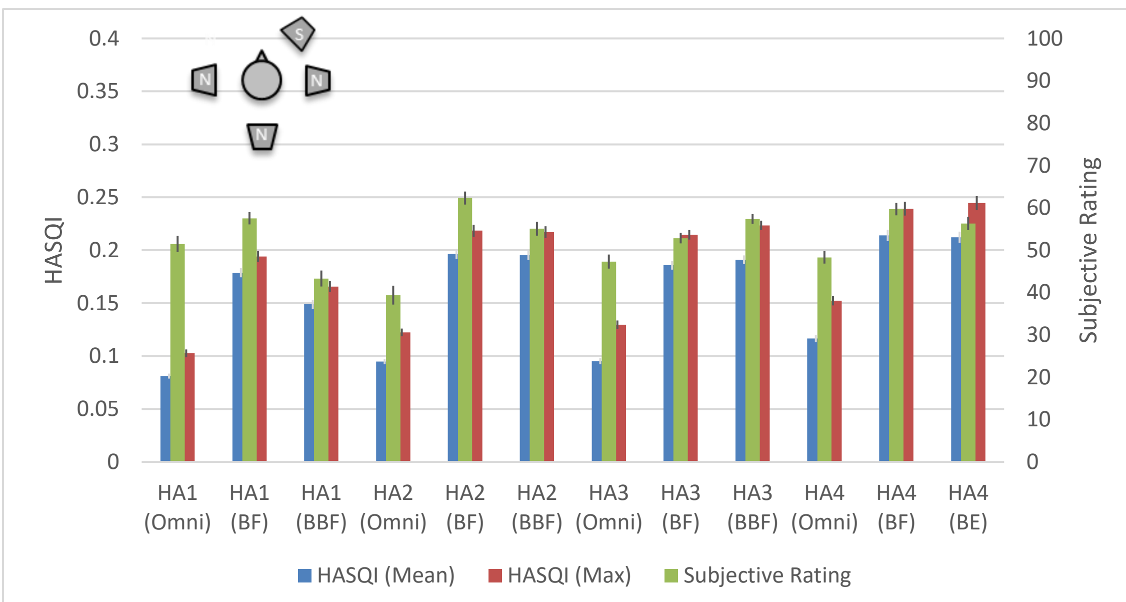


Figure 5.9: Conditionally averaged subjective rating, left-right maximum HASQI score and left-right mean HASQI score of speech in cafeteria noise at 0 dB SNR, with speech originating from 45° and the noise from 90°, 180°, and 270° in the sound booth.

Despite this, the intra-hearing aid patterns were maintained between the subjective for the other directionality conditions as seen in Figure 5.10 and Figure 5.11.

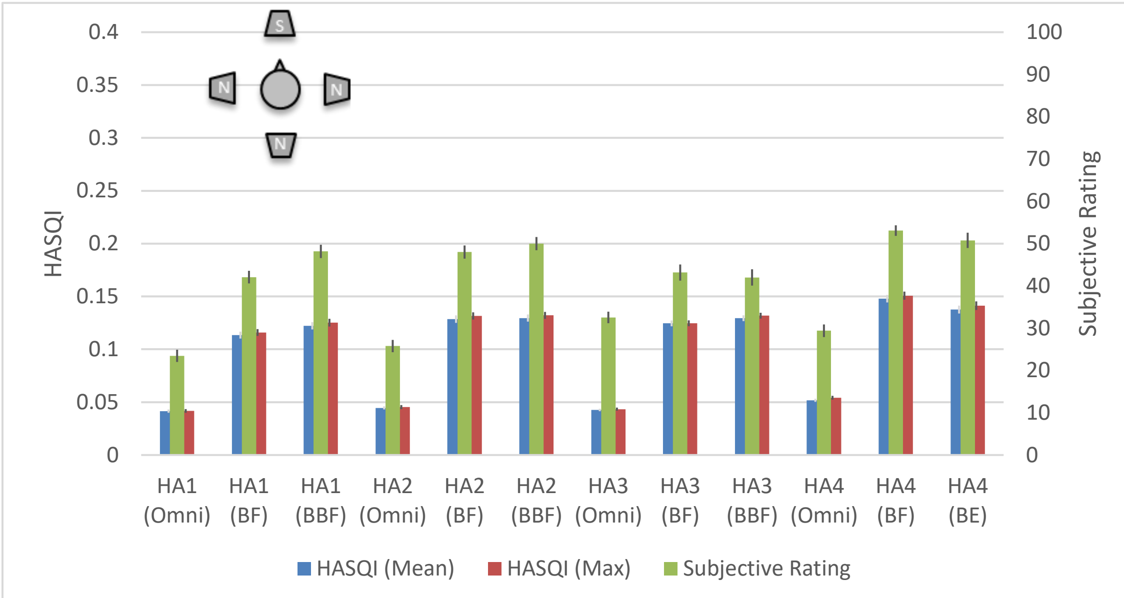


Figure 5.10: Conditionally averaged subjective rating, left-right maximum HASQI score and left-right mean HASQI score of speech in cafeteria noise at 0 dB SNR, with speech originating from 0° and the noise from 90°, 180°, and 270° in the sound booth.

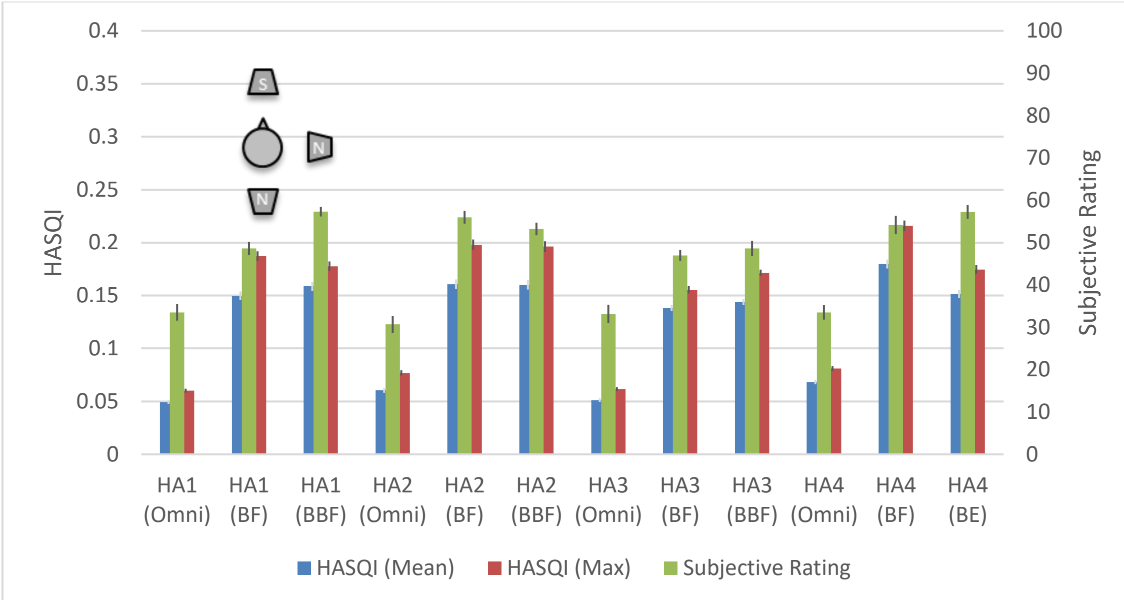


Figure 5.11: Conditionally averaged subjective rating, left-right maximum HASQI score and left-right mean HASQI score of speech in cafeteria noise at 0 dB SNR, with speech originating from 0° and the noise from 90° and 180° in the sound booth.

5.2.2 Reverberation Chamber

Performing the same tests in the reverberation chamber allowed for an analysis into the effect of reverberation on speech quality as well as conditions where the noise was spatially asymmetric but perceptually symmetric due to the diffusion of sound.

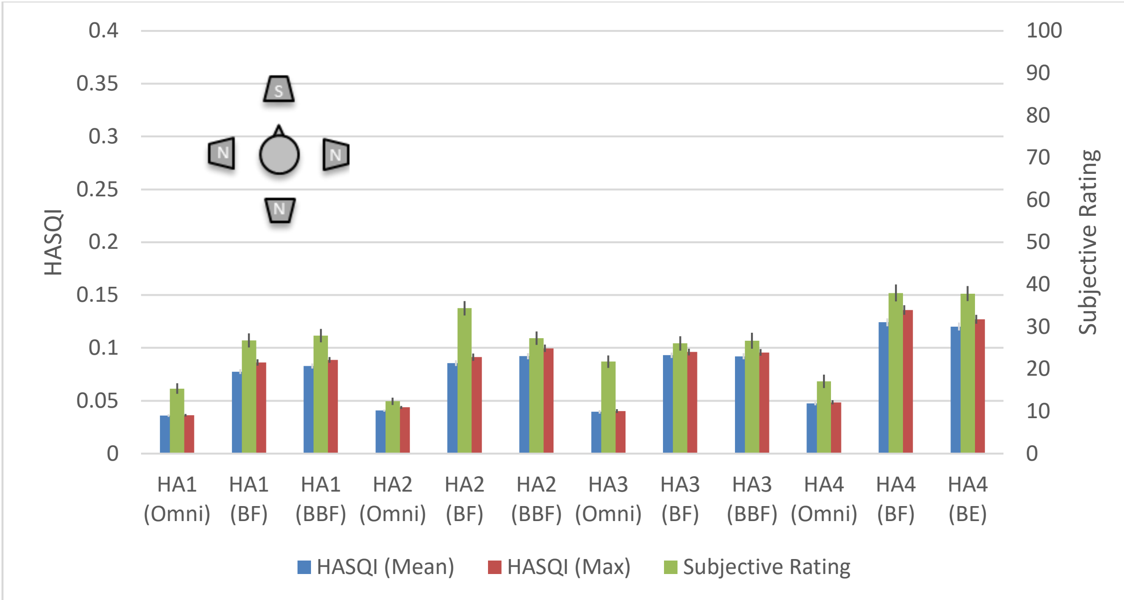


Figure 5.12: Conditionally averaged subjective rating, left-right maximum HASQI score and left-right mean HASQI score of speech in speech-shaped noise at 0 dB SNR, with speech originating from 0° and the noise from 90°, 180°, and 270° in the reverberation chamber.

When the recording is made with noise sources located at 90°, 180° and 270° and the speech source is at 0° in the reverberation chamber, as seen in Figure 5.12, the HASQI scores and subjective ratings largely line up with the electroacoustic analysis where there is an increase in subjective speech quality rating and predicted speech quality from the monaural beamformer to the binaural beamformer in Hearing Aids 1 and 2, and none in Hearing Aid 3. The exception to this is Hearing Aid 2, where the subjective rating for the monaural beamformer is much higher than both the predicted speech quality would imply, as well as the subsequent performance of the binaural beamformer both in subjective ratings and HASQI scoring. In general, subjective ratings obtained for recordings in the reverberation chamber have a lower intra-participant reliability as measured with Cronbach’s alpha.

Hearing Aid 4 performs very well in the reverberant environment in both HASQI scoring and subjective rating, achieving the highest score by both metrics for all directionality conditions.

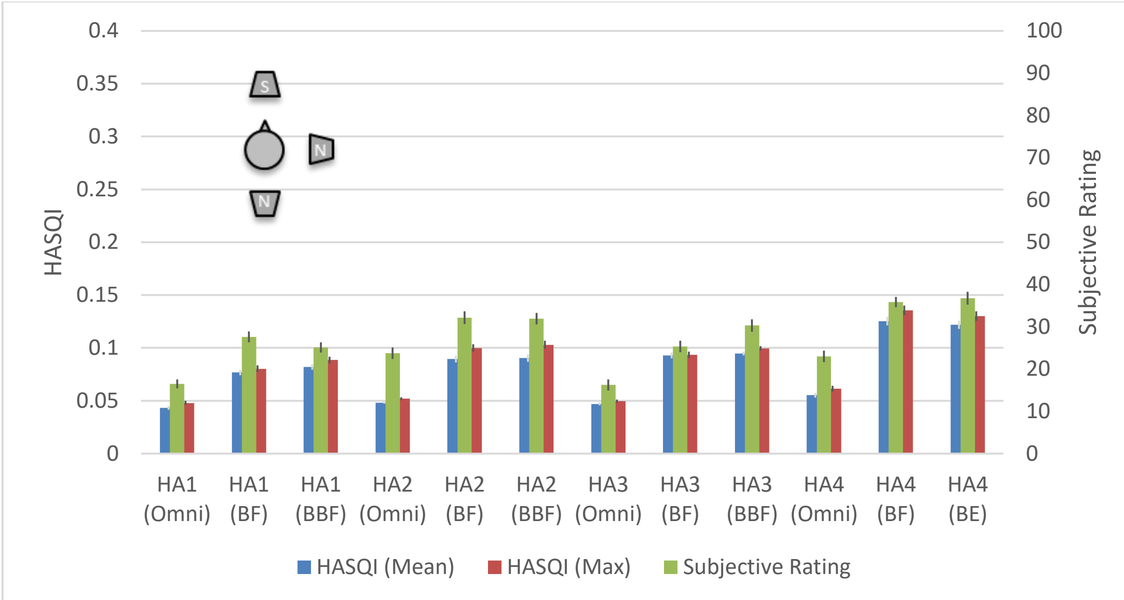


Figure 5.13: Conditionally averaged subjective rating, left-right maximum HASQI score and left-right mean HASQI score of speech in speech-shaped noise at 0 dB SNR, with speech originating from 0° and the noise from 90° and 180° in the reverberation chamber.

When the noise source is located at 90° and 180°, similar results are seen in Figure 5.13 where the HASQI scores are representative of the subjective scores with a few notable exceptions. There are no significant improvements in HASQI scores between the monaural beamformers and the binaural beamformers, however Hearing Aid 1 sees a drop in subjective rating when switched to the binaural beamformer just as Hearing Aid 3 sees an increase.

Discrepancies between the HASQI scores and subjective ratings become apparent when the noise source is located at 90°, 180° and 270° and the speech source is located at 45° as seen in Figure 5.14. The HASQI scores follow the expected outcome when the speech source is outside the range of the narrow binaural beamformer. Hearing Aids 2 and 3 do not improve in predicted speech quality from the monaural beamformer to the binaural beamformer, and Hearing Aid 1 sees a decrease in predicted speech quality. The subjective ratings, however, do not align with these patterns. Both Hearing Aids 1 and 2 see an increase in subjective ratings from the monaural beamformer to the binaural beamformer, and only Hearing Aid 3 sees a decrease in subjective ratings. It becomes clear that the reverberation of the chamber does not affect the participant's subjective rating as much as it does the objective HASQI metric.

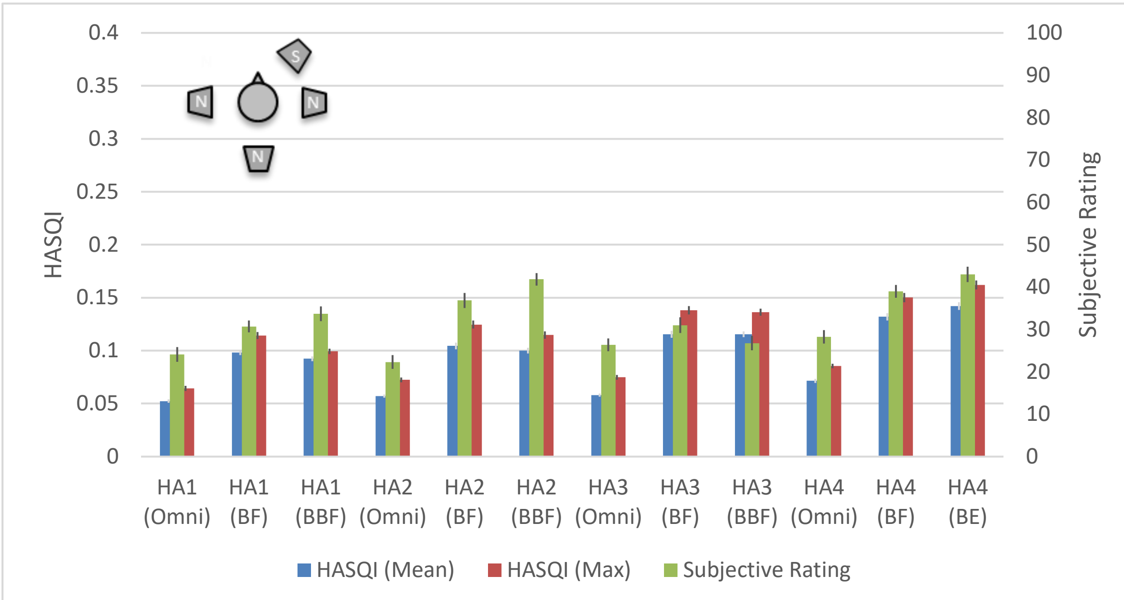


Figure 5.14: Conditionally averaged subjective rating, left-right maximum HASQI score and left-right mean HASQI score of speech in speech-shaped noise at 0 dB SNR, with speech originating from 45° and the noise from 90°, 180° and 270° in the reverberation chamber.

5.3 HASQI Average Weighting Function

By adjusting the weighting function used when averaging the left and right HASQI scores, a stronger correlation between the HASQI average and the subjective ratings can be established. Initially, five HASQI weighting functions were tested, as seen in Table 5.2. The weighting functions take the form of Eq. (5.1).

$$HASQI_{ave} = \mathrm{w}_1 HASQI_{left} + \mathrm{w}_2 HASQI_{right} \tag{5.1}$$

where

$$\mathrm{w}_1 + \mathrm{w}_2 = 1$$

$$1 \geq \mathrm{w}_1, \mathrm{w}_2 \geq 0$$

Table 5.2: HASQI averaging equation comparison.

HASQI Weighting Function	R^2			SSE		
	Sound Booth	Reverb Chamber	Overall	Sound Booth	Reverb Chamber	Overall
Mean	0.9017*	0.7327	**0.7504**	1.473	**1.427**	5.292
Left Ear	0.8438	0.6697	0.7074	1.895	1.791	5.861
Right Ear	0.8645	0.7101*	0.7231	1.647	1.524	5.437
Max	**0.9057**	0.6711	0.7373*	**0.834**	1.728	**4.154**
Min	0.8570	**0.7632**	0.7287	2.708	1.587	7.145
Eqn. (5.5)	0.9004*	0.7037	0.7477*	1.064	1.527	4.607

Bold indicates the best value for that category.
*Indicates value that, when compared to the weighting function with the maximum R squared in that category with Steiger's Z Test [38] [39], has a ρ value less than 0.05 and is therefore not have a statistically significant difference. In other words, the value is statistically similar to the highest R squared value in that category.

The HASQI mean is the average of left and right HASQI scores, and is displayed in the first row of Table 5.2. Weighting the average asymmetrically so it is comprised of entirely the left or right value presents the next two rows on Table 5.2 respectively. Finally, two averaging equations were developed by basing them off of intelligibility weighting tests in literature such as [40] where the ear with the higher intelligibility score drives the intelligibility score up. In other words, the ear with the more intelligible input signal is the most heavily weighted when determining the overall intelligibility. To see if a similar effect presented itself in quality tests, the Max and Min are comprised entirely of the higher HASQI score and the lower HASQI score of the two ears respectively.

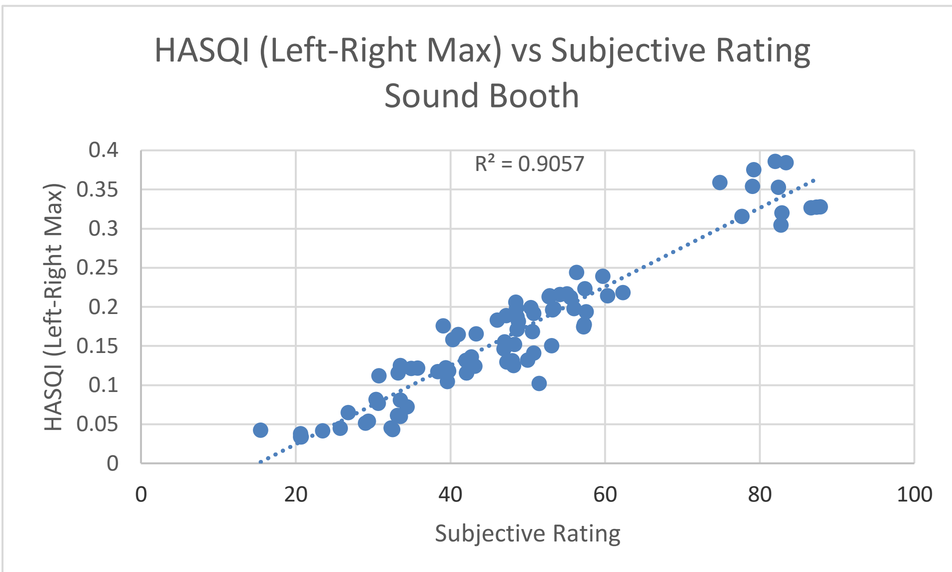


Figure 5.15: Correlation of the conditional averages of the Weighting Function 4 average HASQI score comprising of the maximum HASQI score between the left and right ears and the corresponding recordings subjective ranking for recordings made in sound booth.

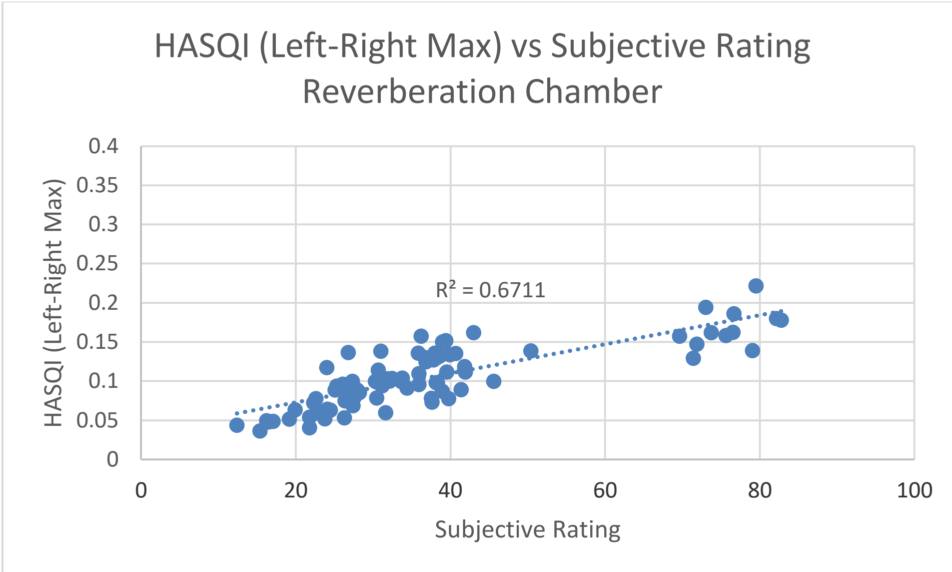


Figure 5.16: Correlation of the conditional averages of the Weighting Function 4 average HASQI score comprising of the maximum HASQI score between the left and right ears and the corresponding recordings subjective ranking for recordings made in reverberation chamber.

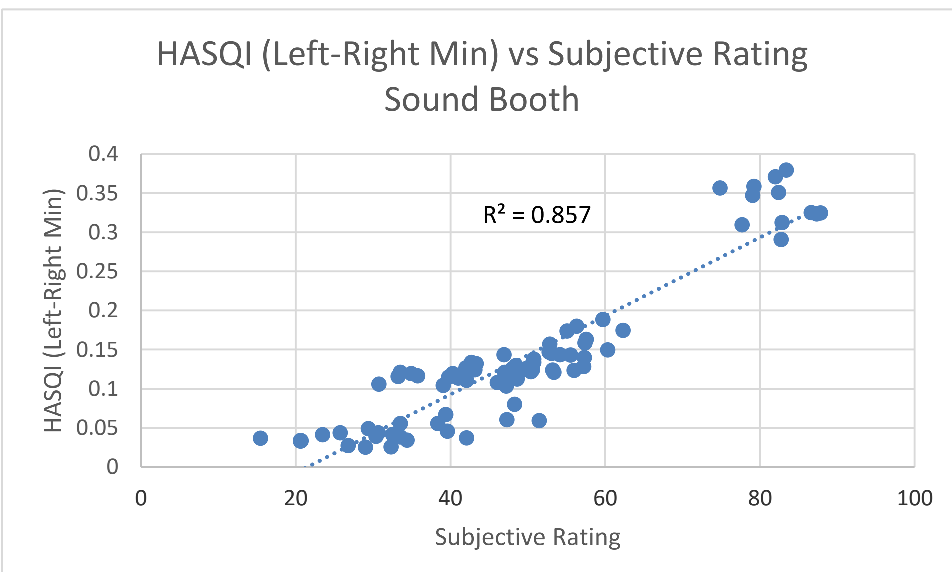


Figure 5.17: Correlation of the conditional averages of the Weighting Function 5 average HASQI score comprising of the minimum HASQI score between the left and right ears and the corresponding recordings subjective ranking for recordings made in sound booth.

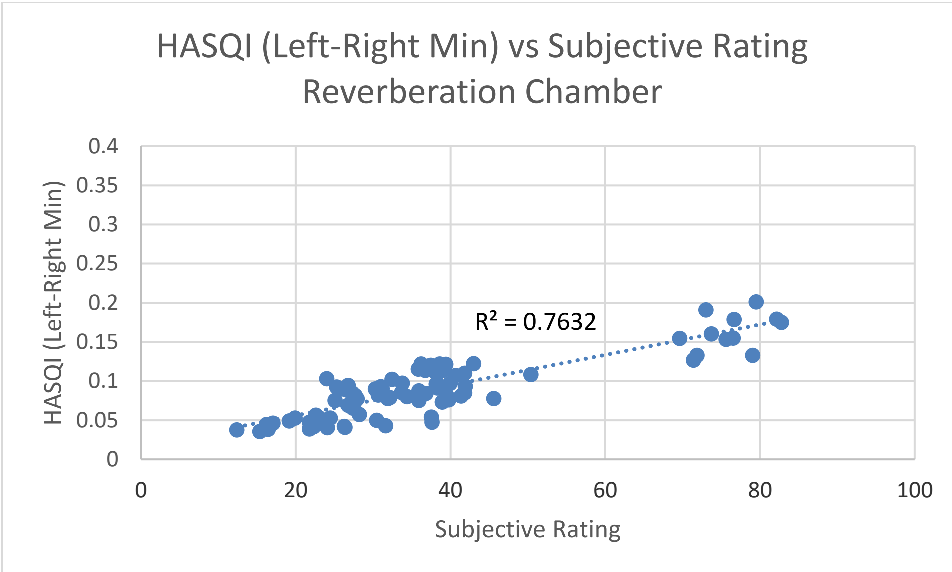


Figure 5.18: Correlation of the conditional averages of the Weighting Function 5 average HASQI score comprising of the minimum HASQI score between the left and right ears and the corresponding recordings subjective ranking for recordings made in reverberation chamber.

By comparing the resulting R squared values of the line of best fit for each average equation versus the corresponding subjective ratings, it can be determined that HASQI maximum has the best fit between the datasets for recordings made in the sound booth while the HASQI minimum has the best fit between datasets for the recordings made in the reverberation chamber. Furthermore, the HASQI mean has the best R squared value overall. This implies at least that the HASQI averaging equation that is the best predictor of overall perceived quality will utilize maximum and minimum HASQI scores, taking the form of Eq. (5.2) seen below. This implication is corroborated by the sum of the squared residual errors (SSE), the lowest of which occur with the maximum HASQI score for recordings made in the sound booth and overall, while the mean HASQI score has the lowest SSE for recordings made in the reverberation chamber.

$$HASQI_{ave} = \mathrm{w}_1 \max(HASQI_{left}, HASQI_{right}) + \mathrm{w}_2 \min(HASQI_{left}, HASQI_{right}) \tag{5.2}$$

where

$$\mathrm{w}_1 + \mathrm{w}_2 = 1$$

$$1 \geq \mathrm{w}_1, \mathrm{w}_2 \geq 0$$

The values of the two weighting coefficient w_1 and w_2 were then calculated using a linear least squares solver in MATLAB with the constraints outlined in Eq. (5.2). The data was first normalized and then used to train the solver to obtain the result in Eq. (5.3).

$$w_1 = 1 \tag{5.3}$$

$$w_2 = 0$$

$$HASQI_{lsq} = \max(HASQI_{left}, HASQI_{right})$$

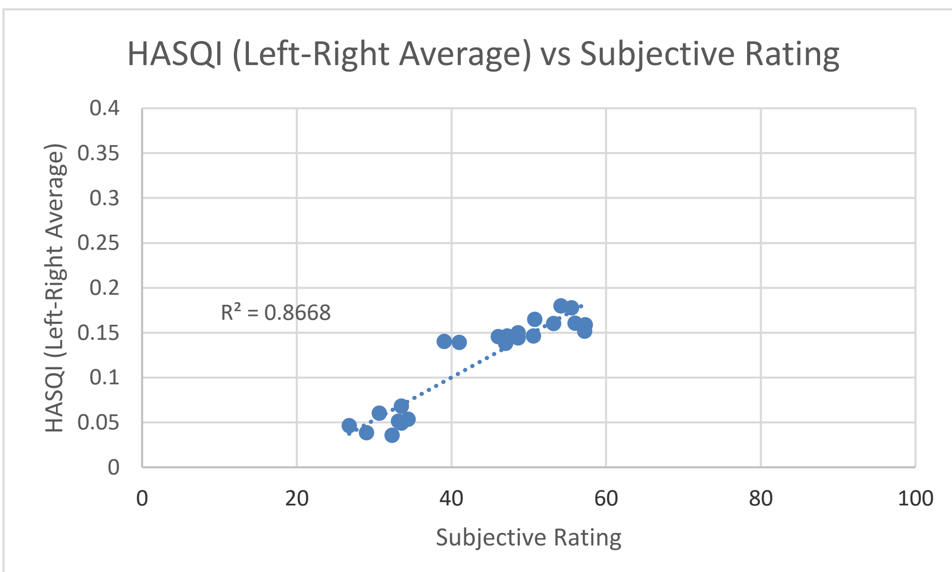
By validating the linear least-squares solver with 10-fold cross validation a correlation of 0.8600 is found. Therefore, according to the linear least squares solution in Eq. (5.3), the HASQI maximum in Table 5.2 is the best fit between the left and right HASQI scores and the subjective ratings when all data is used. When only recordings made in the sound booth are used, the linear least squares solution is Eq. (5.3) again, this time with a correlation of 0.9472. However when only recordings made in the reverberation chamber are used, the linear least squares solution is found in Eq. (5.4). This is much closer to the mean HASQI score.

$$w_1 = 0.4630 \tag{5.4}$$

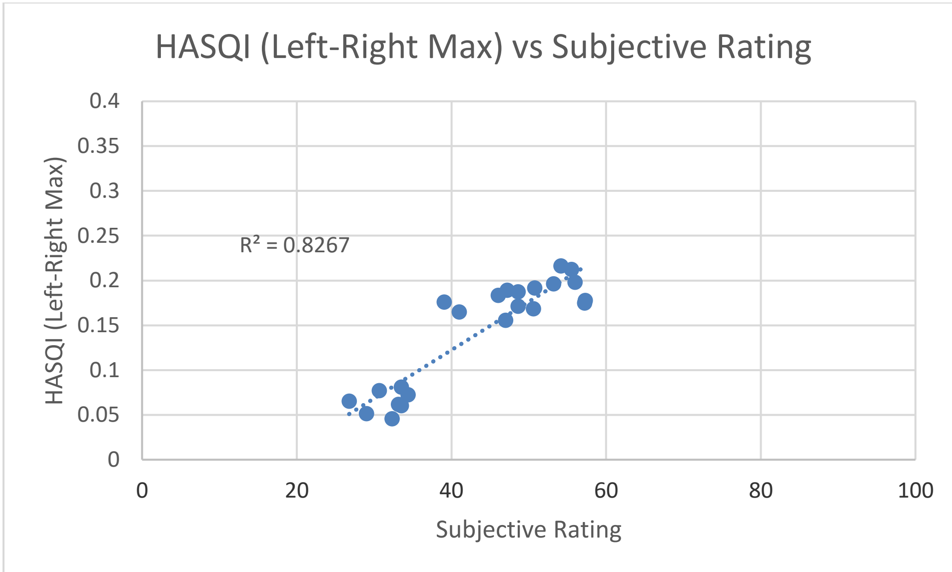
$$w_2 = 0.5370$$

$$HASQI_{reverb} = \mathrm{w}_1 \max(HASQI_{left}, HASQI_{right}) + \mathrm{w}_2 \min(HASQI_{left}, HASQI_{right})$$

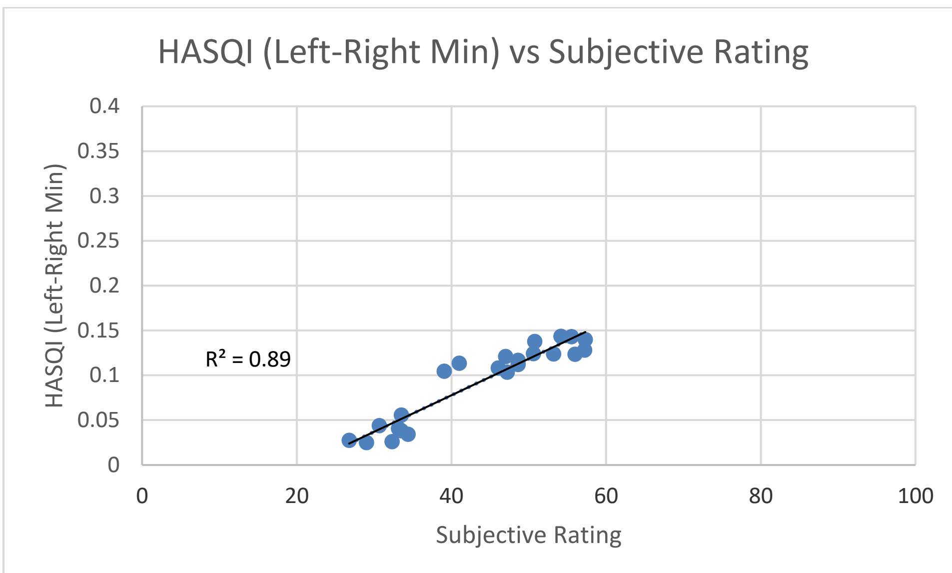
The result in Eq. (5.3) stating that the optimal weighting function for recordings made in the sound booth imply that the subjective rating will generally follow the same pattern as the maximum HASQI score. However as seen in Section 5.2 this is not always the case, particularly in cases with asymmetrical directionality conditions. By restricting the data points to the recordings made in the sound booth in conditions where the noise source is at 90° and 180°, this assumption is verified as seen in Figure 5.19, where the left-right minimum performs the best of the three (Table 5.2: 1, 4, 5) weighting functions similar to scores for recordings made in the reverberation chamber.



a)



b)

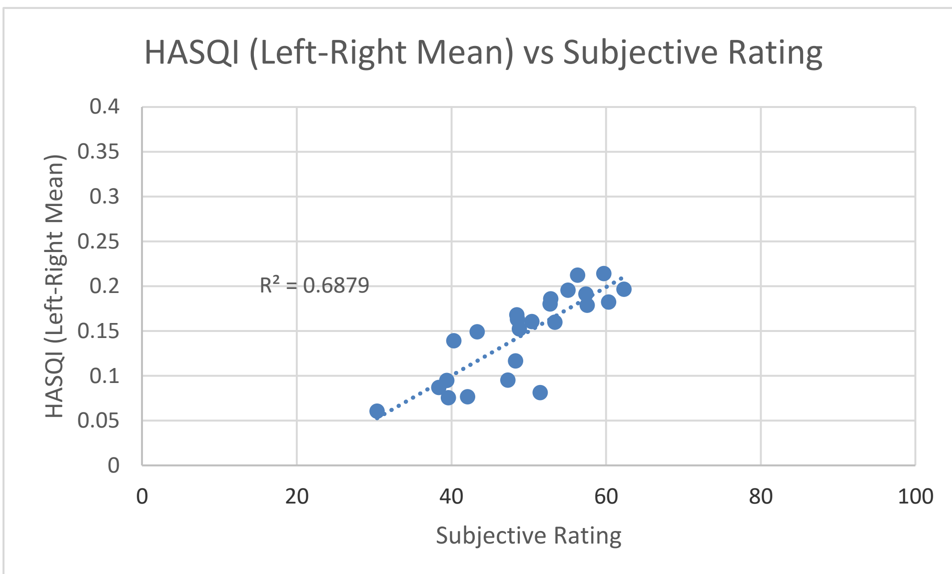


c)

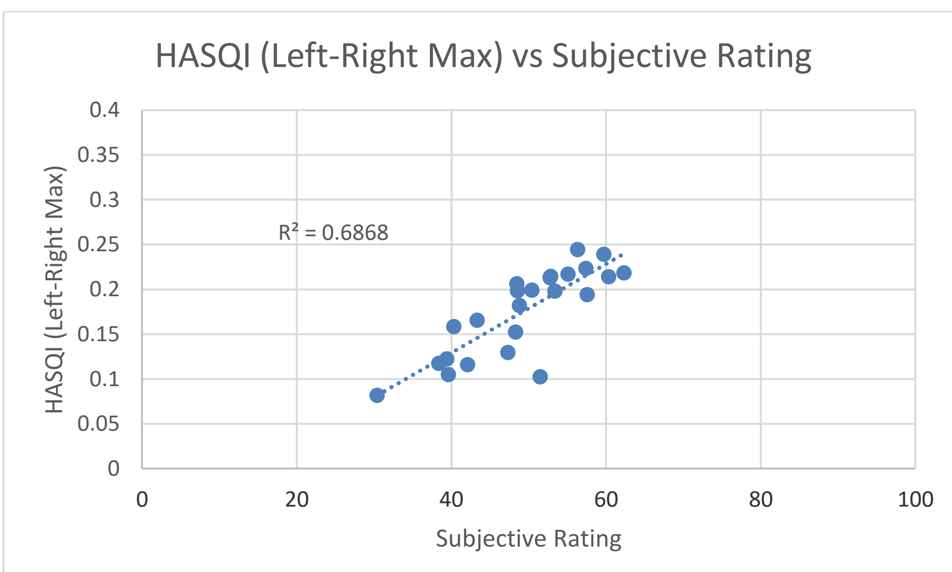
Figure 5.19: Correlation of the conditional averages of a) the left-right mean HASQI score b) the left-right maximum HASQI score c) the left-right minimum HASQI score and the corresponding recording's subjective ranking only for recordings made in the sound booth in conditions where the noise source is at 90° and 180°.

So while overall, the left-right maximum HASQI score is a better correlator of subjective ratings in the sound booth, in cases with asymmetric noise, the left-right minimum HASQI score performs better.

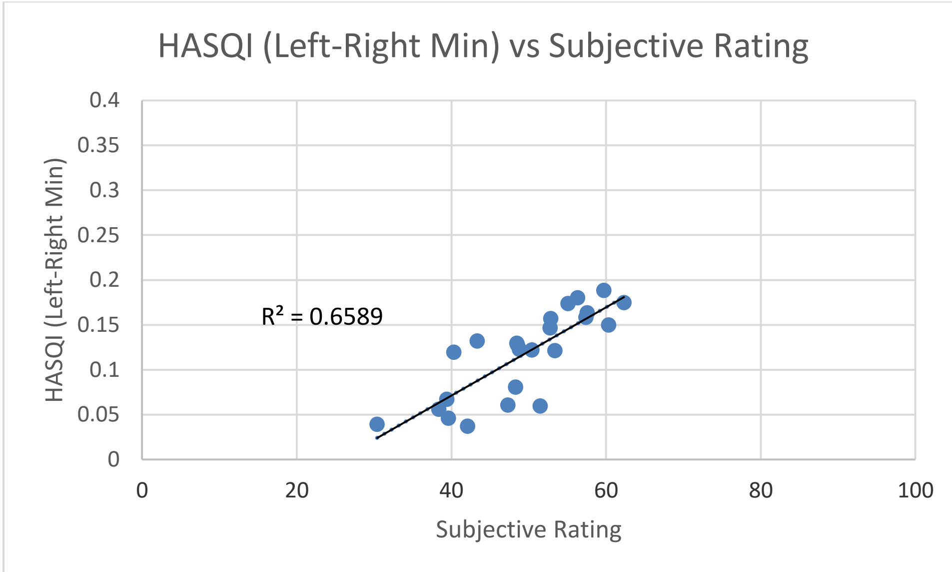
The same analysis can be done by restricting the data points to the recordings made in the sound booth in conditions where the speech source is at 45°, another asymmetrical directionality condition. Unlike the directionality condition where the speech source was located at a 0° angle and the noise sources were located at 90° and 180°, subjective rating correlated better with the HASQI score when it was weighted equally between the left and right ear, as seen in Figure 5.20. Because the best correlation does not occur at either the maximum or minimum extreme, another linear regression analysis must be used to find the best weighting in these asymmetric conditions.



a)



b)



c)

Figure 5.20: Correlation of the conditional averages of a) the left-right mean HASQI score b) the left-right maximum HASQI score c) the left-right minimum HASQI score and the corresponding recording's subjective ranking only for recordings made in the sound booth in conditions where the speech source is at 45° and the noise source is at 90°, 180° and 270°.

To perform the linear regression, data was restricted to only asymmetrical directionality conditions, where the speech source was located at 0° and the noise sources originated from 90° and 180° or where the speech source was located at 45° and the noise sources originated from 90°, 180° and 270°. The data was also limited to recordings made in the sound booth, again to heighten the asymmetry between the left and right scores. The data was then normalized and put through a MATLAB linear least-squares solver with the constraints shown in Eq. (5.3). Using 5-fold cross validation, the resulting model coefficients were averaged and the result is shown in Eq. (5.5).

$$w_1 = 0.7620 \tag{5.5}$$

$$w_2 = 0.2380$$

$$HASQI_{lsq,lmt} = \mathrm{w}_1 \max(HASQI_{left}, HASQI_{right}) + \mathrm{w}_2 \min(HASQI_{left}, HASQI_{right})$$

The result of this solver is less extreme than the previous one. After 5-fold cross validation, the average correlation with the restricted dataset HASQI average with the subjective ratings was 0.8836. While when fitting against all data, Eq. (5.5) has a lower R squared value than the maximum HASQI average found in Eq. (5.3). However comparing the results with Steiger's Z Test [38][39], the ρ-value calculated with a 2-tailed test was found to be less than 0.05 and therefore the difference was not found to be statistically significant.

When restricting the test data set to recordings made in the reverberation chamber and comparing it to the left-right minimum, however, the equation found in Eq. (5.5) was found to have a significantly lower R squared value. With that said, the SSE of Eq. (5.5) was the second lowest of the averaging equations that utilized the maximum and minimum HASQI scores. The performance of the weighting function overall was most indicative of its application where the reverberation of the environment would be unknown.

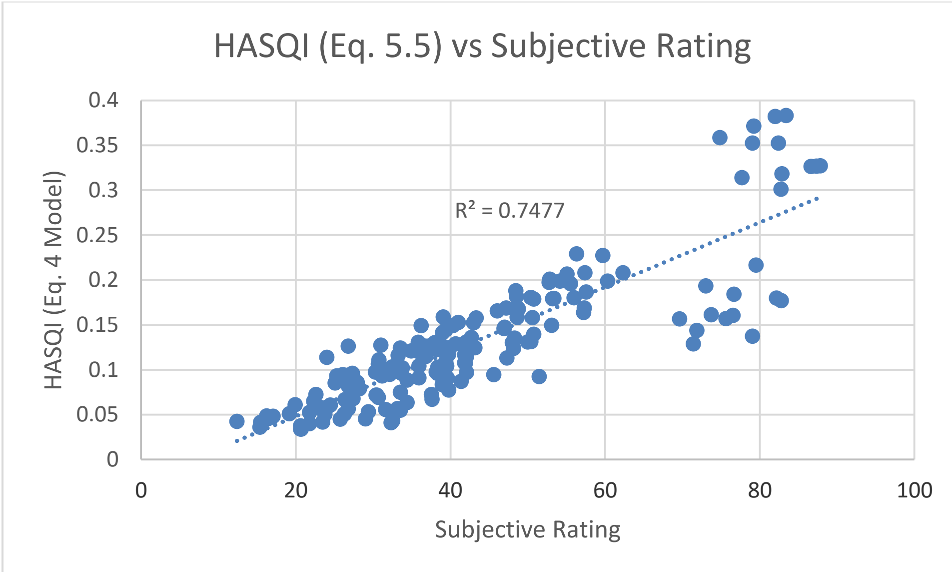


Figure 5.21: Correlation of the conditional averages of Error! Reference source not ound. **model HASQI score and the corresponding recording's subjective ranking.**

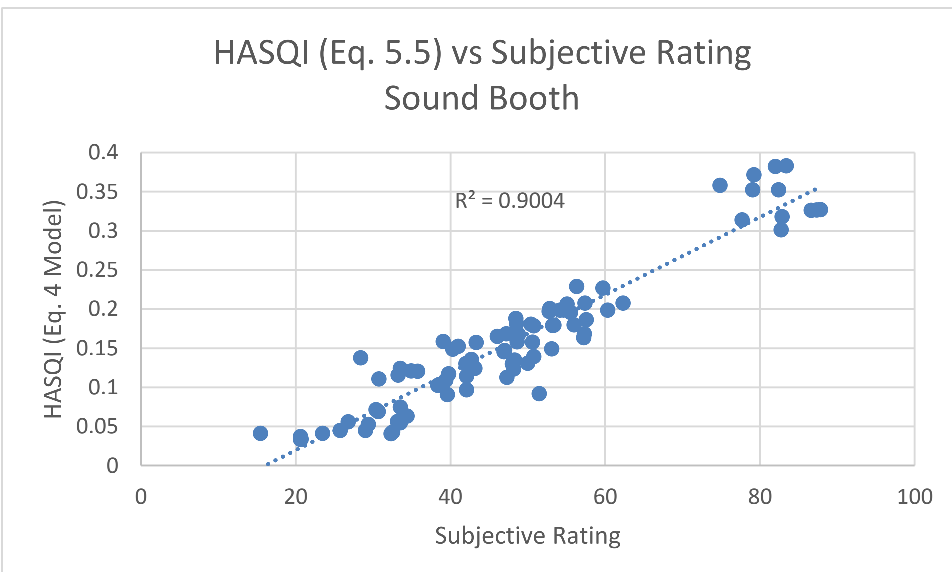


Figure 5.22: Correlation of the conditional averages of (5.5) model HASQI score and the corresponding recording's subjective ranking for recordings made in the sound booth.

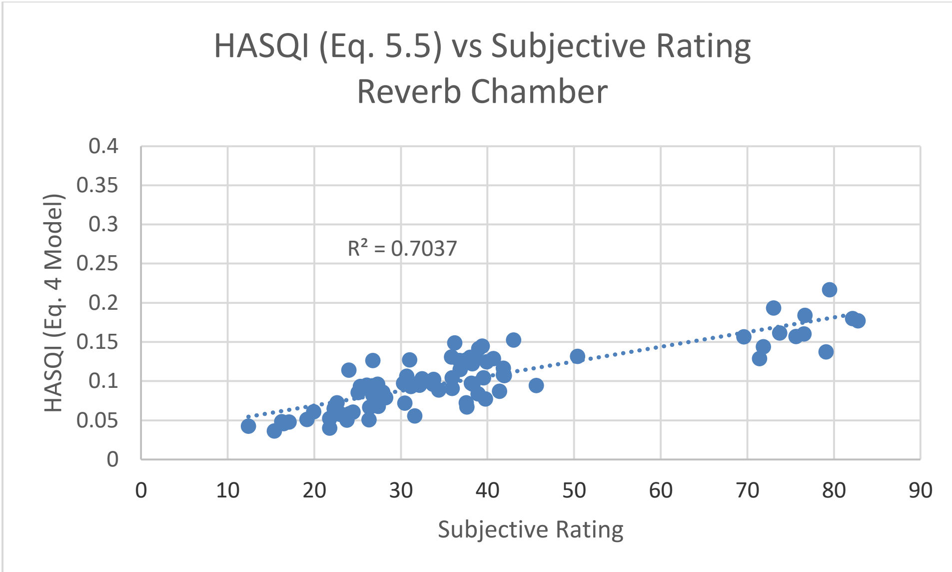


Figure 5.23: Correlation of the conditional averages of (5.5) model HASQI score and the corresponding recording's subjective ranking for recordings made in the reverb chamber.

As Figure 5.19 and Figure 5.20 imply a difference in optimal HASQI Weighting Function between the condition where speech source is at 45°, and the noise source is at 90°, 180° and 270°, and when speech source is at 0° and the noise source is at 90° and 180°, verifying the optimal weighting function for both data sets independently was important. The weighting functions are seen in (5.6) and (5.7) below for the speech at 45° and noise at 90°, 180° and 270°, and the speech at 0° and the noise at 90° and 180° respectively. (5.6) was found to have a 5-fold cross validation correlation of 0.9143 and (5.7) was found to have a 5-fold cross validation correlation of 0.8222.

$$w_1 = 0.6665 \tag{5.6}$$

$$w_2 = 0.3335$$

$$HASQI_{asym1} = \mathrm{w}_1 \max(HASQI_{left}, HASQI_{right}) + \mathrm{w}_2 \min(HASQI_{left}, HASQI_{right})$$

$$\mathrm{w}_1 = 0.7592 \tag{5.7}$$

$$w_2 = 0.2408$$

$$HASQI_{asym2} = \mathrm{w}_1 \max(HASQI_{left}, HASQI_{right}) + \mathrm{w}_2 \min(HASQI_{left}, HASQI_{right})$$

5.3.1 Summary

Analyzing the subjective ratings of the binaural beamforming algorithms provides a more concrete understanding of binaural beamformer's effect on quality. Similar to the electroacoustic results, the results were variable depending on the direction of the speech and noise, the reverberation, and the choice of hearing aid and program. With a few outliers, both monaural and binaural beamformers performed better than the omnidirectional program in all conditions. Additionally, binaural beamformers performed moderately better than monaural beamformers in conditions where speech was located at a 0° angle. When the speech was at a 45° angle, the performance of the binaural beamformers versus the monaural beamformers in subjective ratings dropped for certain hearing aids. Interestingly, in the reverberation chamber, this pattern was not maintained. When speech was at a 45° angle, there was no drop in subjective rating from the monaural beamformer to the binaural beamformer yet certain hearing aids reported a lower objective rating for the binaural beamformer versus the monaural beamformer.

Correlation between the subjective ratings and HASQI scores were fairly high, with an R squared of 0.7504 between the conditionally averaged subjective ratings and the mean HASQI score. By optimizing the weighting function used in the average with all data, a lower sum of squared errors can be achieved without significantly affecting the correlation. However as this weighting function has a poor correlation with the asymmetrical directionality conditions, a new weighting function was solved for using only the asymmetrical data in Eq. (5.5). This led to a new R squared of 0.7477 for all data, which has a statistically insignificant difference than the weighting function currently in usage as well as being valid for asymmetrical speech and noise conditions.

Chapter 6

6 Conclusion

The continued development and perfection of hearing aids is important to further enhance the experience of hearing aid users. Part of the constant updating of hearing aid technology involves testing and ensuring new features provide a tangible benefit to users and not just an extra cost. With wireless technology that allows hearing aids to communicate with each other becoming more ubiquitous, it is important to ensure that the benefits the binaural connection provides to noise reducing beamforming algorithms is quantifiable in a perceptually relevant manner. Not only that, but ensuring there are strategies and processes in place to quickly and effectively test these algorithms with objective metrics allows more developments in the technology to be made with ease.

6.1 Goals

This thesis sought to meet three goals. First, benchmarking binaural beamforming algorithms with electroacoustic intelligibility metrics, electroacoustic quality metrics, and behavioural quality metrics. Salient results from this thesis relevant to the first goal are:

- Monaural and binaural beamformers generally perform better than omnidirectional programs in predicted speech intelligibility and quality in electroacoustic tests.

- Binaural beamformers perform slightly better than monaural beamformers in conditions where the speech source is at a 0° angle and noise is surrounding the user in the sound booth. Benefits were reduced when noise was asymmetrical, and binaural beamformers often performed worse in electroacoustic tests when the speech source was at a 45° angle.

- Reverberation affected the performance of the hearing aids, but binaural beamformers still performed better in electroacoustic tests in conditions where the speech source is at a 0 angle°.

- Monaural and binaural beamformers perform better than omnidirectional programs in subjective quality ratings.

- Binaural beamformers generally performed better than monaural beamformers in subjective quality ratings, even when the speech source was at a 45° angle.

- A lot of variation in performance remained between hearing aid models in both electroacoustic and behavioural tests.

The second goal was to validate HASQI with subjective quality ratings when binaural beamforming was activated. As performing electroacoustic quality tests with HASQI is much easier logistically than behavioural tests, ensuring the scores gleaned from HASQI are representative of subjective quality in cases with binaural beamforming algorithms was important for the endorsement of HASQI for future test cases.

- The mean HASQI score correlated well with the subjective ratings for recordings made in the sound booth.

- The mean HASQI score did not correlate well with the subjective ratings for recordings made in the reverberation chamber. However HASQI scores maintained similar patterns between recordings made in the sound booth and reverberation chamber.

The final goal was the development of a weighting function that could combine left and right HASQI scores in a perceptually relevant way. There are many test cases where upon switching a program, the HASQI score in one ear may rise while the HASQI score in the other drops. It can be difficult to compare test cases in such a scenario without a single index. Therefore, finding a weighting function which could harness the relationship between the two ears for quality was an important step in developing long-term procedures which could be used to electroacoustically benchmark binaural hearing aids.

- Optimizing the weighting function for all the data led to a weighting function that used only the larger of the left and right HASQI scores, however this weighting function did not correlate well with recordings made in asymmetrical speech noise conditions.

- Optimizing the weighting function for only the data collected in the sound booth, where there was a high correlation with the subjective ratings, and in asymmetrical speech and noise conditions, led to a weighting function that gave the largest of the left and right HASQI scores more weight but still had a component of the smaller score. This weighting function correlated well with recordings made in asymmetrical speech and noise conditions and still performed statistically similar to the mean HASQI score with the rest of the data.

6.2 Future Work

As binaural beamformers are a relatively new technology, additional steps can be taken to develop further testing procedures.

- As recordings made in the reverberation chamber did not correlate as well with the subjective ratings, a further investigation of the effect of reverberation on binaural beamformers and sound quality as predicted by HASQI would be beneficial, including recordings made in environments with varying amounts of reverberation.

- Recordings made with higher SNRs would also be beneficial to add more data points which fall in the mid to high range of HASQI.

- The performance of the binaural beamformers was highly dependent on the direction of the speech source. However the performance was not always consistent between brands. Therefore, an investigation on the performance of binaural beamformers in speech directions from 0° to 90° to determine how the direction of the speech source affects the predicted speech quality and intelligibility would be valuable for determining the exact range of speech source angles the binaural beamformer remains beneficial to the user.

References

[1] H. Dillon, *Hearing Aids*. New York, Sydney: Boomerang Press, 2001.

[2] G. R. Popelka, B. C. J. Moore, R. R. Fay, and A. N. Popper, *Hearing aids*. Springer, 2016.

[3] A. W. Bronkhorst, "The cocktail-party problem revisited: early processing and selection of multi-talker speech," *Attention, Perception, Psychophys.*, vol. 77, no. 5, pp. 1465–1487, 2015.

[4] K. Strom, "A Marketing Perspective on Phonak's New Quest," *Hearing Review*, 2013. [Online]. Available: http://www.hearingreview.com/2013/03/a-marketing-perspective-on-phonak-s-new-quest/.

[5] H. Kamkar-Parsi, E. Fischer, and M. Aubreville, "New Binaural Strategies for Enhanced Hearing," *Hearing Review*, 2014. [Online]. Available: http://www.hearingreview.com/2014/10/new-binaural-strategies-enhanced-hearing/.

[6] M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.*, 1994.

[7] ITU-R BS.1534-3, "Method for the subjective assessment of intermediate quality level of audio systems," *Int. Telecommun. Union*, 2015.

[8] J. M. Kates and K. H. Arehart, "The hearing-aid speech perception index (HASPI)," *Speech Commun.*, 2014.

[9] J. M. Kates and K. H. Arehart, "The Hearing-Aid Speech Quality Index (HASQI)," *J. Audio Eng. Soc.*, vol. 58, no. 5, pp. 363–381, 2010.

[10] J. M. Kates and K. H. Arehart, "The Hearing-Aid Speech Quality Index (HASQI) Version 2," *J. Audio Eng. Soc.*, vol. 62, no. 3, pp. 99–117, 2014.

[11] R. M. Cox, G. C. Alexander, and C. Gilmore, "Development of the Connected Speech Test (CST)," *Ear Hear.*, vol. 8, no. SUPPLEMENT, p. 119s, Oct. 1987.

[12] R. Beutelmann and T. Brand, "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.*, 2006.

[13] N. R. French and J. C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," *J. Acoust. Soc. Am.*, 1947.

[14] B. W. Y. Hornsby, "The speech intelligibility index: What is it and what's it good for?," *Hear. J.*, vol. 57, no. 10, pp. 10–17, 2004.

[15] L. D. Kryter, "Validation of the Articulation Index," 1962.

[16] H. J. M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.*, 1980.

[17] L. E. Humes, D. D. Dirks, T. S. Bell, C. Ahlstbom, and G. E. Kincaid, “Application of the Articulation Index and the Speech Transmission Index to the Recognition of Speech by Normal-Hearing and Hearing-Impaired Listeners,” *J. Speech Lang. Hear. Res.*, 1986.

[18] A. N. Standard, “Methods for Calculation of the Speech Intelligibility Index,” *ANSI S3.5-1997*, 1986.

[19] J. Ma, Y. Hu, and P. C. Loizou, “Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions,” *J. Acoust. Soc. Am.*, vol. 125, no. 5, p. 3387, 2009.

[20] K. S. Rhebergen, N. J. Versfeld, and W. A. Dreschler, “Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise,” *J. Acoust. Soc. Am.*, vol. 120, no. 6, pp. 3988–3997, 2006.

[21] R. Beutelmann and T. Brand, “Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.*, vol. 120, no. 1, pp. 331–342, 2006.

[22] J. M. Kates and K. H. Arehart, “Coherence and the speech intelligibility index,” *J. Acoust. Soc. Am.*, 2005.

[23] C. H. Taal, R. C. Hendriks, R. Heusdens, J. Jensen, and U. Kjems, “An evaluation of objective quality measures for speech intelligibility prediction,” in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2009.

[24] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, “An algorithm for intelligibility prediction of time-frequency weighted noisy speech,” *IEEE Trans. Audio, Speech Lang. Process.*, 2011.

[25] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, “Perceptual Evaluation of Speech Quality (PESQ) - A new method for speech quality assessment of telephone networks and codecs,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001.

[26] R. Huber and B. Kollmeier, “PEMO-Q - A new method for objective audio quality assessment using a model of auditory perception,” *IEEE Trans. Audio, Speech Lang. Process.*, 2006.

[27] T. Rohdenburg, V. Hohmann, and B. Kollmeier, “Objective perceptual quality measures for the evaluation of noise reduction schemes,” in *9th International Workshop on Acoustic Echo and Noise Control*, 2005.

[28] J. Kates, “An auditory model for intelligibility and quality predictions,” vol. 050184, no. 2013, pp. 050184–050184, 2013.

[29] A. A. Kressner, D. V. Anderson, and C. J. Rozell, “Robustness of HASQI,” *IEEE Trans. Audio, Speech Lang. Process.*, 2013.

[30] E. F. H. Kamkar-Parsi M. Aubreville, “Improving speech\intelligibility in noise with binaural beamforming,” *Siemens Audiol.*, 2014.

[31] T. Rohdenburg, V. Hohmann, and B. Kollmeier, “Robustness analysis of binaural hearing aid beamformer algorithms by means of objective perceptual quality measures,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2007.

[32] R. O. Duda and W. L. Martens, “Range dependence of the response of a spherical head model,” *J. Acoust. Soc. Am.*, vol. 104, no. 5, pp. 3048–3058, 1998.

[33] J. Attias, E. M. Picou, E. Aspell, and T. A. Ricketts, “Potential benefits and limitations of three types of directional processing in hearing aids,” *Ear Hear.*, 2014.

[34] E. M. Picou, T. M. Moore, and T. A. Ricketts, “The Effects of Directional Processing on Objective and Subjective Listening Effort,” *J. Speech Lang. Hear. Res.*, 2017.

[35] G. Kidd, C. R. Mason, V. Best, and J. Swaminathan, “Benefits of acoustic beamforming for solving the cocktail party problem,” *Trends Hear.*, 2015.

[36] E. M. Picou and T. A. Ricketts, “An Evaluation of Hearing Aid Beamforming Microphone Arrays in a Noisy Laboratory Setting,” *J. Am. Acad. Audiol.*, vol. 30, no. 2, pp. 131–144, 2018.

[37] N. Bisgaard, M. S. M. G. Vlaming, and M. Dahlquist, “Standard Audiograms for the IEC 60118-15 Measurement Procedure,” *Trends Amplif.*, vol. 14, no. 2, pp. 113–120, 2010.

[38] J. H. Steiger, “Tests for comparing elements of a correlation matrix,” *Psychol. Bull.*, vol. 87, no. 2, pp. 245–251, Mar. 1980.

[39] I. A. Lee and K. J. Preacher, “Calculation for the test of the difference between two dependent correlations with one variable in common,” *Computer Software*. 2013.

[40] B. A. Edmonds and J. F. Culling, “The spatial unmasking of speech: Evidence for better-ear listening,” *J. Acoust. Soc. Am.*, vol. 120, no. 3, pp. 1539–1545, 2006.

Appendices

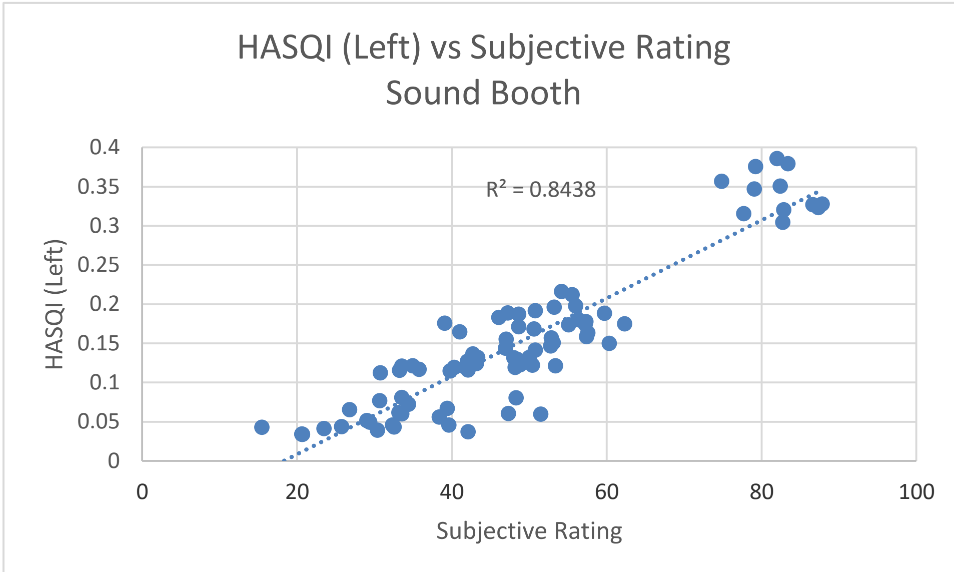
Appendix A: Subjective test participant information.

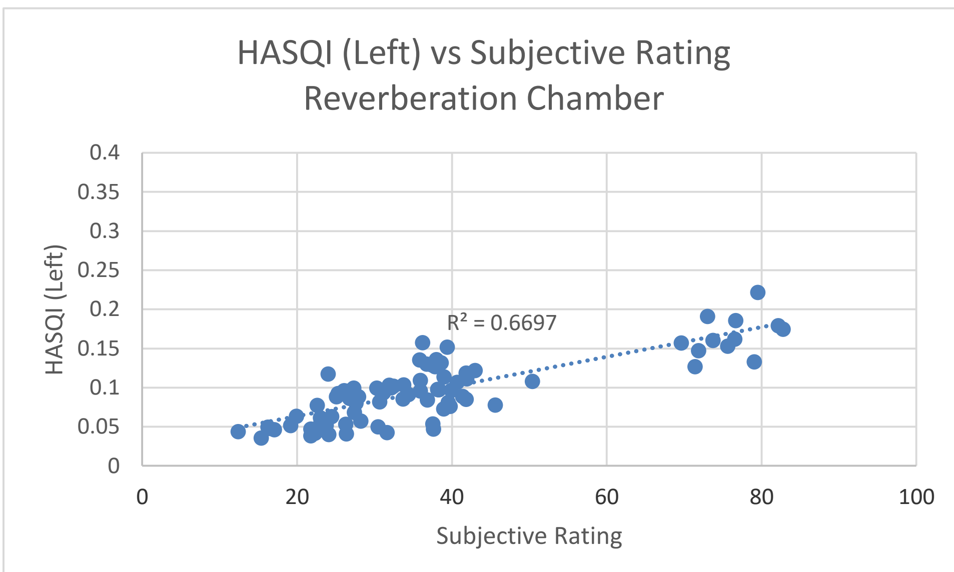
Participant	Age
1	82
2	72
3	78
4	72
5	69
6	76
7	69
8	68
9	73
10	82
11	60
12	76
13	67

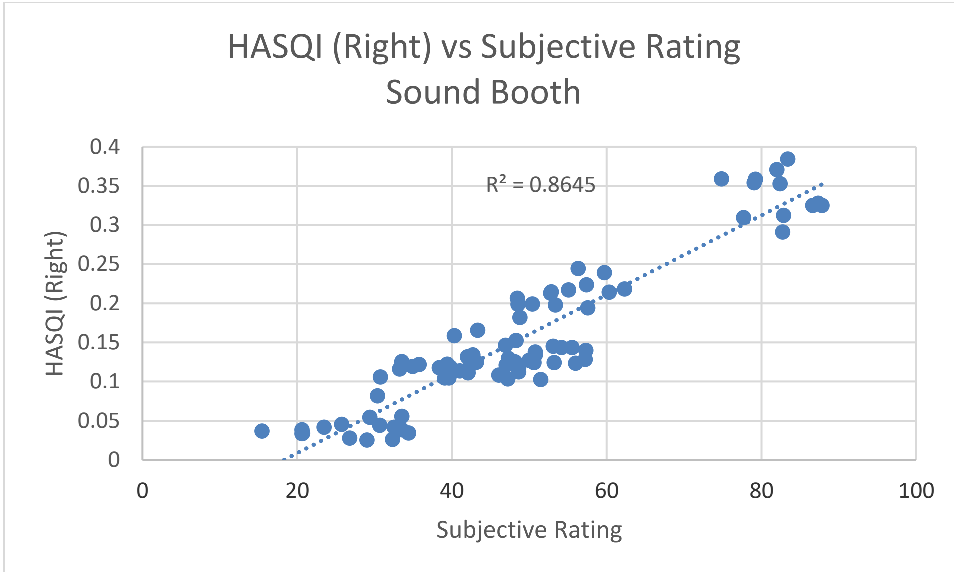
Appendix B: KEMAR RECD values.

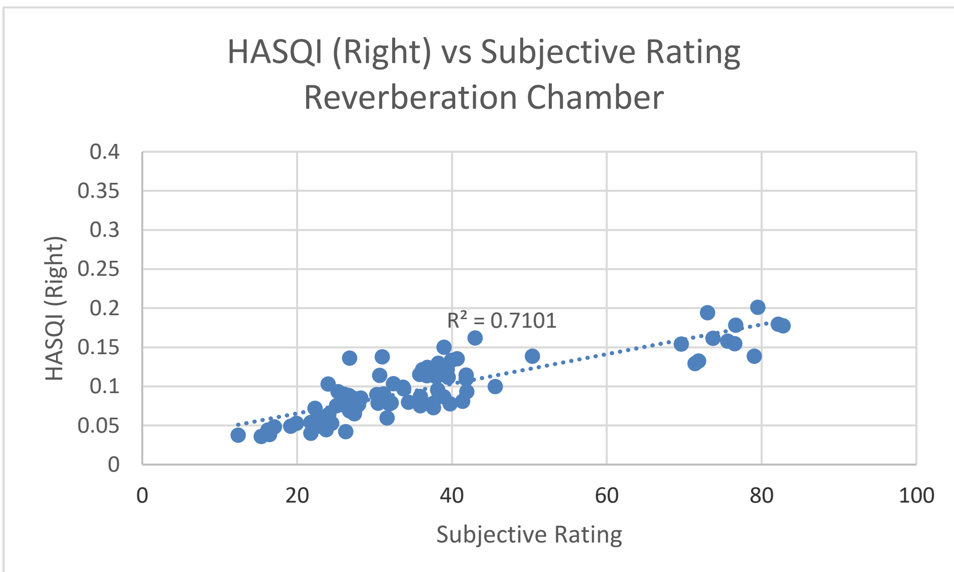
250	500	1000	2000	4000	6000
2.9	3.6	5.6	9.2	15	18

Appendix C: Correlation of the conditional averages and the corresponding recordings subjective ranking.






HASQI (Right) vs Subjective Rating
Sound Booth
HASQI (Right)
Subjective Rating
R² = 0.8645
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
20
40
60
80
100


HASQI (Right) vs Subjective Rating
Reverberation Chamber
HASQI (Right)
Subjective Rating
R² = 0.7101
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
20
40
60
80
100

Appendix D: Western University Health Sciences Research Ethics Board approval letter

Western Research

Date: 8 February 2019

To: Dr. Susan Scollie

Project ID: 111665

Study Title: Advanced hearing tests as predictors of hearing aid benefit

Application Type: HSREB Amendment Form

Review Type: Delegated

Meeting Date / Full Board Reporting Date: 26Feb2019

Date Approval Issued: 08/Feb/2019 18:29

REB Approval Expiry Date: 24/Oct/2019

Dear Dr. Susan Scollie ,

The Western University Health Sciences Research Ethics Board (HSREB) has reviewed and approved the WREM application form for the amendment, as of the date noted above.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
Laminated_Gabrielson SQ rating scale	Paper Survey	20/Jan/2019	
Listening Effort Scale	Paper Survey	20/Jan/2019	
Scollie 111665 Letter of 20-01-2019	Consent Form	20/Jan/2019	
Scollie 111665 Protocol _20-01-2019	Protocol	20/Jan/2019	

Documents Acknowledged:

Document Name	Document Type	Document Date	Document Version
Summary of Changes 20_01_2019	Summary of Changes	20/Jan/2019	

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University HSREB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Please do not hesitate to contact us if you have any questions.

Sincerely,

nicola Geoghegan-Morphet, Ethics Officer on behalf of Dr. Philip Jones, HSREB Vice-Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Curriculum Vitae

Name:	Scott Aker
Post-secondary Education and Degrees:	University of Western Ontario London, Ontario, Canada 2013-2017 B.E.Sc.
	University of Western Ontario London, Ontario, Canada 2017-2019 M.E.Sc.
Honours and Awards:	Ontario Graduate Scholarship 2017-2018, 2018-2019
Related Work Experience	Teaching Assistant University of Western Ontario 2017-2019
	Software Intern (DSP) Audioscan 2018

Publications:
S. Cho, S. Aker, and E. Macpherson, "Assessment of a feasible virtual acoustics method for testing hearing aids using the Hearing-Aid Speech Perception Index," *Journal of the Acoustical Society of America*, vol. 145, no. 3, pp. 1880, Apr. 2019.